TOWARD A MATHEMATICAL THEORY OF PERCEPTION

By

James Thomas Kajiya

Technical Report #4116

Department of Computer Science

California Institute of Technology

A dissertation submitted to the faculty of the University of Utah in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Copyright © James T. Kajiya 1979
All Rights Reserved

ABSTRACT

A new technique for the modelling of perceptual systems called formal modelling is developed. This technique begins with qualitative observations about the perceptual system, the so-called perceptual symmetries, to obtain through mathematical analysis certain model structures which may then be calibrated by experiment. The analysis proceeds in two different ways depending upon the choice of linear or nonlinear models. For the linear case, the analysis proceeds through the methods of unitary representation theory. It begins with a unitary group representation on the image space and produces what we have called the fundamental structure theorem. For the nonlinear case, the analysis makes essential use of infinite-dimensional manifold theory. It begins with a Lie group action on an image manifold and produces the fundamental structure formula.

These techniques will be used to study the brightness perception mechanism of the human visual system. Several visual groups are defined and their corresponding structures for visual system models are obtained. A new transform called the Mandala transform will be deduced from a certain visual group and its implications for image

processing will be discussed. Several new phenomena of brightness perception will be presented. New facts about the Mach band illusion along with new adaptation phenomena will be presented. Also a new visual illusion will be presented. A visual model based on the above techniques will be presented. It will also be shown how use of statistical estimation theory can be made in the study of contrast adaptation. Furthermore, a mathematical interpretation of unconscious inference and a simple explanation of the Tolhurst effect without mutual channel inhibition will be given. Finally, image processing algorithms suggested by the model will be used to process a real-world image for enhancement and for "form" and texture extraction.

CONTENTS

ACKNOWLEDGEMENTS

There are many that should be thanked here but space permits mention of only the outstanding few. First I would like to express gratitude to the members of my committee Elaine Cohen, John Smith and Thomas G. Stockham, Jr. Special thanks go to Professor Stockham without whose advice and guidance the quality of this work might have been considerably poorer.

For hours of technical discussions I would like to thank Tracy Petersen and Jim Youngberg. For support and technical assistance thanks go to Steve Boll, Brent Baxter, and Mike Milochik. Also I would like to thank Stewart Ogden for his help and advice toward survival in a sometimes hostile bureaucracy.

Finally, I wish to thank my wife, Barbara whose support and encouragement have been essential to this work. In spite of her full time professional duties she has rendered aid both in criticism and in preparation of this final document. Her generous contribution has exceeded the already considerable amount customarily given by a spouse under these circumstances.

CHAPTER 1

INTRODUCTION

1.1 Research Summary

The research reported below takes some first steps toward studying processes that occur within human perceptual systems in a mathematical way. Specifically, we shall present a method we have termed formal modelling for obtaining a series of structure formulas for possible models of perceptual systems. We shall apply this theory to the problem of studying the brightness perception mechanism in human vision, i.e. the study excludes temporal and chromatic considerations. Furthermore several new visual phenomena will be presented. A new visual model which elicits responses to a wide variety of stimuli in a manner similar to the visual system will be discussed. Finally this model will be used to process real-world images in some new ways.

1.2 System Identification

To understand some of the problems encountered in developing models for a perceptual system, say the visual system, we shall discuss some methods of system identification popular in engineering circles and discuss

their applicability to visual modelling.

The classical method of system identification treats the system as a black box with a priori structure. For example, a simple model describing a biological system may be written as

$$x^{k+1} = Ax^k + Bu^k$$

where x is an n-dimensional output vector, u is an m-dimensional input vector and A and B are unknown matrices of the appropriate dimensions. We define a new vector

$$w^k = \begin{pmatrix} x_1^k \\ \vdots \\ x_n^k \\ u_1^k \\ \vdots \\ u_m^k \end{pmatrix}$$

and a new unknown matrix

$$\Phi = [A;B]$$

to obtain equations

$$x^{k+1} = \Phi w^k .$$

The idea is to identify the unknown Φ by observing the input w and output x. We do this by calculating a linear least squares regression estimate $\hat{\Phi}$ of Φ given by (Graupe 1972):

$$\phi_i' = [W_k' W_k]^{-1} W_k' \chi_i^{k+1}$$

where

Φ_i' is the i^{th} row of Φ

$$W_k = \begin{pmatrix} w_1^k \\ \vdots \\ w_r^k \end{pmatrix}$$

w_μ^k is the μ^{th} measurement of w^k

$$\chi_i^{k+1} = \begin{pmatrix} x_{i,1}^{k+1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_{i,r}^{k+r} \end{pmatrix}$$

$x_{i,\mu}^{k+1}$ is the μ^{th} measurement of i^{th} component of x^{k+1}.

Note that the data needed are the matrices W_k, that is, r examples of the input-output behavior of the system. For biological systems the above procedure is inadequate because biological systems are almost always strongly nonlinear. To treat nonlinear systems, we can extend the classical method by augmenting the w^k vector with judiciously chosen nonlinear functions, i.e.,

$$w^k = \begin{pmatrix} x_1^k \\ \vdots \\ x_n^k \\ f_1(x^k) \\ \vdots \\ f_p(x^k) \\ u_1^k \\ \vdots \\ u_m^k \\ g_1(u^k) \\ \vdots \\ g_s(u^k) \end{pmatrix} .$$

The procedure then follows in an identical manner. Hopefully, a felicitous choice of nonlinear functions will allow a small increase in dimensionality to effect a large improvement in fit.

A second method of system identification is the Volterra-Wiener method of nonlinear system identification (Palm and Poggio 1977) which has enjoyed wide popularity in the modelling of various biological systems. Briefly, these methods treat an operator as a parameterized family of functionals. Then each functional is developed as a series of polynomial functionals (see Hille and Phillips 1958), i.e.,

$$S_\chi : u(\xi) \longmapsto \sum_{n=0}^{\infty} K_n u$$

where

$$K_n u = \int \cdots \int k_n(\xi_1, \ldots, \xi_n)\, u(\xi_1)\, u(\xi_2) \ldots u(\xi_n)\, d\xi_1 \ldots d\xi_n.$$

The kernels k_n are then found by observing the input-output behavior of a system S on a white noise test signal. Let $u(\xi)$ be the white noise input, $x(\xi)$ be the output. Then the system kernels k are given by

$$k_n(\xi_1, \ldots, \xi_n) = u(\tau - \xi_1)\, u(\tau - \xi_2) \ldots u(\tau - \xi_n)\, x(\tau)\, d\tau.$$

This method has enjoyed a wide popularity among those charged with modelling biological systems. There are a number of characteristics of this method worth mentioning. First, the need to guess the augmented nonlinearities in the previous method is avoided. Second, computational effort is reduced. Lastly, because this method restricts the system estimate to polynomial functionals, delicate considerations regarding convergence must be met. In practice, if convergence is not attained in two terms, the method becomes unwieldy.

A third method of system identification occurs as a part of the blind deconvolution process (Stockham et al. 1975). In this process, a source signal s(t) has been degraded by an unknown linear time invariant system with impulse response h(t) and the corrupted output signal $v(t)=s(t)*h(t)$ is observed. The problem is to estimate the impulse response h(t), that is, identify the system, then construct an appropriate "inverse" kernel to compute a restoration of v(t). Only the system identification process will interest us here.

In the system identification step for blind deconvolution, a broad assumption is posited with regard to the second-order statistics of the input signal. Roughly the observed signal v(t) is segmented into finite length records $v_i(t)$ whose size is hopefully longer than the impulse response h(t) of the degrading system. The cepstra $\hat{v}_i(\lambda)$ of $v_i(t)$ are taken and then averaged. We have

$$\hat{v}_i(\lambda)=\hat{s}_i(\lambda)+\hat{h}_i(\lambda)\approx\hat{s}_i(\lambda)+\hat{h}(\lambda).$$

The broad assumption is that the average of $\hat{s}_i(\lambda)$ converges to some known prototype $\bar{s}(\lambda)$

$$\frac{1}{N}\sum_{i=1}^{N}\hat{s}_i(\lambda)\approx\bar{s}(\lambda).$$

This prototype is then subtracted from the sample mean

$$\frac{1}{N}\sum_{i=1}^{N}\hat{v}_i(\lambda)$$

to obtain $\hat{h}(\lambda)$.

The plausibility of applying all three of the above techniques to perceptual modelling suffers from three requirements shared by the methods to varying degrees. The first requirement is that the experimenter must be able to precisely control any aspects of the input signal to which the visual system may be sensitive. The second requirement is an ability to precisely measure the output signal of the visual system. The third requirement is the need for an accurate system postulate. Let us examine the ramifications of these three requirements.

To precisely control any aspects of the input signal

to which the visual system may be sensitive is, as most psychophysicists know, an expensive and painstaking task. Of the three methods for system identification discussed above, the first is relatively lenient with regard to this requirement while the second and third require a precise control of second-order statistics of the input signal.

The second requirement is an ability to measure the output signal of the visual system. Even though psychophysics has been remarkably successful in obtaining quantitative measurements of various operating parameters of the visual system, an observation of the output signal of the visual system as a whole seems impossible. All of the above system identification methods require detailed observations of such a hypothetical signal. Thus their direct application to visual modelling is hampered severely by this requirement.

It is the last requirement, for a precise a priori system structure, that is the greatest weakness of the first method. The second and third methods limit the system structure to polynomial and linear time-invariant operators, respectively.

The experimenter, in order to utilize the first technique, must postulate a system model structure and calibrate that model by performing an experiment. If the fit to data is good he declares himself done. If the fit is not good the experimenter has one of two choices: 1) he

can discard the present model and postulate a different system structure or 2) he can add another variable to the present system structure so that the model will fit the data gathered from the experiment just performed. Unfortunately, as computer scientists are especially aware, the human tendency is toward the second path. Thus the postulated model accumulates more and more parameters with each passing experiment until eventually the model becomes so complicated that its behavior is extremely difficult to analyze. The predictive power of such a model then becomes negligible. This second path we have termed ad hoc modelling.

1.3 Symmetries and Structure

In contrast to the above situation we shall present a method which we term formal modelling which is in large measure free from these three requirements. The first and second requirements are circumvented by requiring only a very fundamental observation about the visual system -- the symmetries the visual system obeys. We are able to use only this fundamental observation because the results produced by the formal modelling procedure will be less specific than that of a complete modelling procedure. The techniques discussed below will not seek to produce a specific model for the visual system, but rather a model structure which can then be calibrated by experiment.

Because a fundamental observation forms the premise and a model structure is the final product of formal modelling, the disadvantages of ad hoc modelling are avoided.

The class of transformations of an image that preserve the qualitative aspects of that image embody the symmetries that the visual system obeys. For example, when an image is translated laterally, our visual system has no trouble in recognizing that the image is the same except for translation. However, when one constructs a transformed image by reciprocating each light intensity (as in a photographic negative), significant qualitative shifts occur. Thus the visual system is not symmetric under such an operation.

When one concatenates two symmetry operations in succession, their product is (within limits) a symmetry operation. Similarly, for every symmetry operation there exists another symmetry operation such that the product of the two is the identity operation. These facts allow us to say that the set of symmetry operations forms a group, or more precisely a group germ.

It is this group, along with the representation of this group as a set of operators on the space of all images, which forms the starting point for the theory discussed below. The symmetry group and its representation are the above-mentioned fundamental observations about the visual system.

From this observation we will be able to partition the space of images into certain small building blocks or "bricks" such that the visual system operates very simply on each brick. This process is identical to that of constructing a transform domain in which the visual system may be described as a simple transfer function. This transform domain, or set of bricks, forms the structure which the model must have.

For example, in the classical orthogonal series expansions one develops an input signal into a sum of certain orthogonal basis functions. If those basis functions are eigenvectors for the system that will eventually process such signals, then we may describe the action of such a system by a transfer function, i.e. a function that assigns to each eigenvector the eigenvalue of the system under consideration. Each brick in this case is a single eigenvector and the operation of the system is simply scalar multiplication by the appropriate eigenvalue.

In classical linear system theory, the bricks, or eigenvectors, are the complex exponentials. We then obtain a situation familiar to every engineer: the process of partitioning the space is simply Fourier transformation, the simple description becomes a frequency response. For every frequency we measure the system's response to the complex exponential, that is, we measure

the eigenvalues of the system corresponding to a single complex exponential.

That the complex exponentials may be considered to be eigenfunctions of the visual system explain the importance of the Fourier transform for image processing. This also explains why many other orthogonal expansions are less useful for the processing of images.

What the symmetry theory does is to, in a generalized sense, calculate the eigenvectors of the class of systems symmetric under a given group. Formal modelling hinges on the remarkable fact that with knowledge of only the symmetry group one can calculate the eigenfunctions of the system -- without any further information. That is, the two parts of the system, its eigenfunctions and its eigenvalues can be determined separately: the eigenfunctions follow from the symmetries, the eigenvalues characterize the specific system in the symmetry class.. Notice the Fourier analog. The Fourier trnasform fixes the eigenvectors while the frequency response, i.e. the eigenvalues, characterizes a specific system symmetric under shifts. Thus with a knowledge of only the symmetry group of the system we wish to consider, we will be able to construct a generalized sort of transform into a domain which is specifically tailored to represent the action of that system in a simple way. This is what we mean by formal modelling.

1.4 Preview

The following provides a brief summary of the subsequent chapters. For the reader interested in only the basic idea, most of the technical mathematics can be avoided by reading sections 2.3, 3.2 and 4.2 which deal with only special cases of the theory.

Chapter 2 provides some definitions and examples for subsequent chapters. Chapter 3 develops the theory of group representations as far as needed in order to develop a structure theory for linear operators. Chapter 4 presents results that form a structure theory for nonlinear operators. Chapter 5 applies the theories developed in the previous two chapters to a specific group to obtain a psychophysically interesting transform -- the Mandala transform. Chapter 6 discusses some well known visual brightness perception phenomena and presents some new visual phenomena. Chapter 6 also discusses some well established brightness perception models and evaluates their performance with respect to the phenomena discussed. A new model is presented in Chapter 7 along with a few comments upon the implications that this model has with regard to the image processing discipline. Furthermore, two image-processing experiments using a real-world image will be presented. Finally, conclusions and suggestions for further research are given in Chapter 8.

CHAPTER 2

PRELIMINARIES

2.1 Perceptual Symmetries and Groups

The techniques of group theory and the study of vision are linked via the concept of perceptual symmetry. Symmetry in the way we shall use it is to be interpreted in a related but more general sense than the immediate geometric notion. In geometry one usually states that a figure possesses symmetry if under a specified class of geometric transformations that figure is taken into itself, e.g. a figure is said to possess bilateral symmetry if the transformation takes each half into its mirror image; similarly, a circle is symmetric under all rotations about its center. Our concept of symmetry will be similar to the concept of kinematic symmetry in physics. Under this more general concept of symmetry, the behavior of the system is not so much unchanged by a given class of transformations as it is modified in some particularly graceful way. A system obeying Newton's laws of motion is kinematically symmetric under Galilean transformations, i.e. the time evolution operator is symmetric under such transformations. If the initial

positions and velocities of the system are transformed then the time evolution operator does not produce subsequent motion identical to the untransformed case but rather subsequent motion also modified by the same Galilean transformation.

A similar example is a visual symmetry which could be called size change. When an image is scaled such that each linear extent is modified to the same degree we say that the image has undergone a dilatation. As an image is scaled our perception of it varies in a particularly graceful way; our general impression remains the same save a particular aspect which our language terms size. This transformation of the image enjoys a status shared by only a very limited subset of the possible transformations that could be applied to images.

Care should be taken to avoid confusing the notions of symmetry and invariance. The latter implies some quantity unchanging under an allowed set of transformations while the former implies some restricted type of change. The notion of invariance manifests itself as the phenomenon of perceptual constancies. For example, size constancy is the tendency for the visual system to judge the size of an object independently of the actual angular extent of such an object's projection upon the retina. A pencil whose distance from the eye is doubled has the size of its projection upon the retina halved, yet

we do not see it so. These perceptual constancies have been long studies by psychophysicists and their relation to perceptual symmetries is close. The relation is so close, in fact, that confusion of the two is quite common in the literature. We claim that it is important and, as the remaining chapters hope to show, most fruitful to keep the distinction between these two concepts clear.

Cassirer (1944) was the first to recognize that the perceptual transformations, viz. transformations under which the visual system is symmetric, obey the mathematical axioms for a group. After some analysis he concluded that the visual system uses this fact to form certain "perceptual universals". Later, Pitts and McCulloch (1947) conjectured that the brain actually performs a crude sort of harmonic analysis to construct a set of invariances which are the basis for such universals. They reasoned that the brain may extract the "DC" component, i.e. the group average, of an image to obtain an invariant under the symmetry group considered. This work has been followed up by Dunn (1973,1975) to construct line and corner extractors. The work in this report also performs a harmonic analysis, but for more than the DC component. Hoffman (1966,1971) considered the many possible symmetry groups for vision but since his contribution was oriented toward the geometric illusions (line drawings instead of images) he did not consider any

harmonic analysis but constructed various differential invariants on the visual manifold.

We shall now present a number of groups we feel are important to the study of vision. Completeness is not the intent of the following list. Rather it is hoped that the search for new, more powerful visual groups will continue so that in the future the formal modelling procedure may be reapplied with greater result.

The first group we discuss is the translation group, which we write as VG_0. This group is isomorphic to $(R^2,+)$. Certainly an image that has been translated remains subjectively unchanged in almost all respects. Perhaps translation could be called the strongest perceptual symmetry. These comments apply, of course, only to translation within moderate bounds. This issue will be taken up following discussion of all the groups. It is interesting to note that application of the formal modelling procedure to the translation group leads naturally to the Fourier transform domain. This does much toward explicating the special status of the Fourier transform with regard to other orthogonal function expansions. When considering a specific class of visual behavior, different symmetry groups may vary in their applicability. The translation group has the widest, though not universal, applicability of any of the groups in this list.

We shall in the sequel be frequently concerned with one-dimensional images, viz. images which have no luminance variation in one (usually vertical) spatial dimension. These images are quite familiar to anyone even casually acquainted with the psychophysical literature. We write VG_1 for the group of dilatation and translation of such images, that is, images such that the x-axis undergoes a possible scale change and/or translation. This group is also known as the ax+b group. Application of this group to the study of form perception is readily seen once one considers the phenomenon that the form of an image is not seen to change with distance from that image. On the other hand, when it is desired to measure the detectability of one-dimensional gratings, scale becomes a critical parameter. This suggests that VG_1 may not be useful in the detection paradigm (although, as we shall see, this suggestion is misleading). For the overall brightness perception system, it follows from an early observation by Stockham (see Baudelaire (1973, p. 55)) that VG_1 is a symmetry group if the relative perceived brightness rather than absolute brightness is considered. Thus if the components of the visual system responsible for detection are not to be independent of the rest of the brightness perception system, VG_1 will be useful in the study of the detection system also.

When considering the full class of two-dimensional

images we will have in mind the group VG_2 of translations in the plane and dilatations. This group is similar to VG_1 except that an additional degree of freedom has been added in the translations. Incidentally, it is widely held that dilatation symmetry is the second most powerful visual symmetry (translation being the first) known. Our ability to recognize instantly and without scrutiny scaled versions of some configuration is strong evidence for such a belief (see Julesz (1971, Section 3.3)).

Another visual symmetry group VG involves rotations and translations. Here the application is not so much to form perception as to detection and brightness perception studies. The inapplicability of this group to form perception stems from many experiments. As an example we discuss the results of Shepard and Metzler (1973). They found that recognition time of images of three-dimensional blocklike constructions varies proportionally to the degree of rotation, irrespective of the axis of rotation (a rotation in or out of the plane of the paper). Such a linear relationship does not occur between recognition time and degree of translation or dilatation -- the recognition is virtually instantaneous. A full discussion of orientation and form perception can be found in Rock (1973). The case for applicability of VG_3 to detection and brightness perception studies is easy to make, however, since neither the detectability nor the

perception of brightness of a stimulus is significantly altered by action of VG_3.

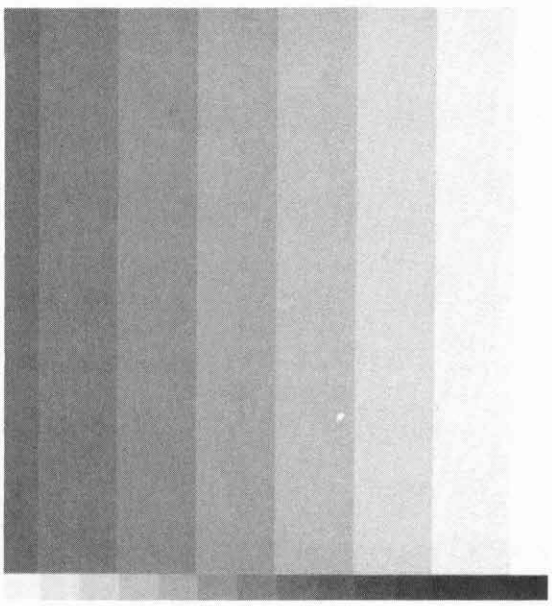
The group VG_4 will be the combination of VG_2 and VG_3. Again the argument that various components of the visual system possess symmetries in VG_2 and VG_3 indicates that a combination of the two may be useful in probing the structure of the visual system.

The last group we consider is VG_6 which we shall call the gamma group. The action of this group is the operation of taking a point-by-point γ-power of the image, that is,

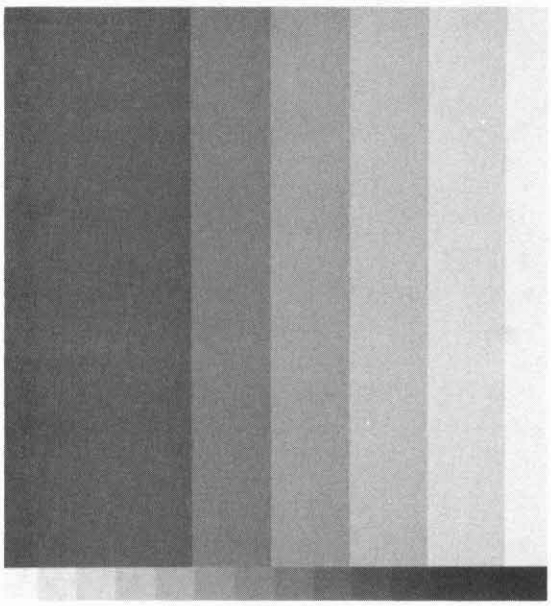
$$f(x,y) \longmapsto \{f(x,y)\}^{\gamma}$$

for $\gamma>0$. This group is isomorphic to the multiplicative group of positive reals. Figure 2.1 displays an image with a number of transformations. Note that the gamma operation maintains the relationship of the brightness of the components. Photographers are well aware of the fact that gamma is a symmetry in vision. The applicability of this group to form and brightness perception is obvious.

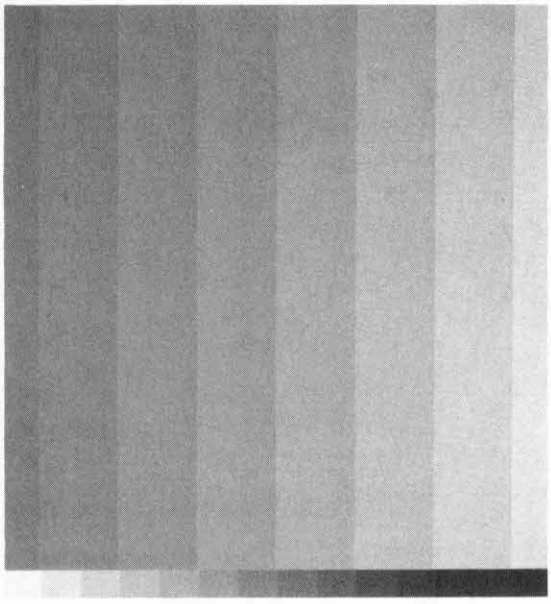
We shall now discuss the bounds within which the above groups may act upon the visual manifold. We observe that most of the symmetries considered apply to the visual system within limits. For example, dilatation is certainly not a symmetry at extreme magnifications or radical reductions but must rather be limited to a moderate scale change. These comments apply to



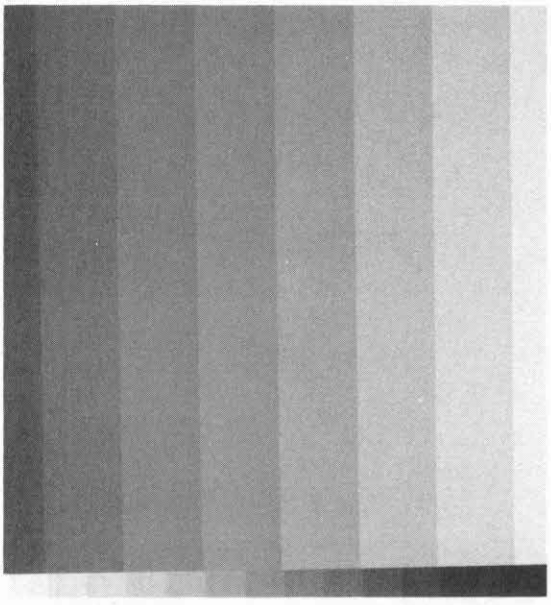
(a)



(b)



(c)



(d)

Figure 2.1 - Transformations of an image. (a) Original (b) Bias (c) Gamma Power (d) Constant multiple

translations and the gamma group as well. Thus we should strictly be avoiding the use of the full group and concentrate on the group germ, i.e. the portion of the group in the neighborhood of the identity. We can therefore consider the use of these groups as a simplifying approximation to the more complicated groups defined on a visual manifold more sophisticated than R^2. Hoffman (1966) gives some symmetry groups based on the Luneburg binocular visual space. It should be noted that these more complicated groups possess the same germs as in our simplified list. Furthermore, the methods presented in Chapter 4 make use of only these germs.

The reader may have gotten the feeling that the action of VG_5 is significantly different from those of the previous groups. This is indeed so in two important ways. First, the action of the previous groups is principally expressed in terms of the underlying visual manifold. The corresponding transformation on images is such that the operator would be said by engineers to possess memory. Second, the action of VG_5 upon images, although memoryless, is nonlinear. It is most doubtful that any but the most trivial linear systems would be symmetric under such a group, and indeed the visual system is nonlinear. Therefore a linear symmetry theory will be inadequate to the task of analyzing systems symmetric under such groups. We shall study the linear theory

nonetheless because it sheds important light upon linear approximations to the visual system (such as the detection paradigm perturbation approximation) and is most useful when coupled with methods for considering nonlinear systems in a linear framework (the homomorphic theory). Furthermore, Chapter 4 will present a nonlinear symmetry theory which directly contributes to the formal modelling process.

2.2 Basic Definitions

A topological group is a group such that the point set of elements in G carry a topology. This topology is required to be fine enough to make the group operations continuous. We will usually require the topology of G to be locally compact so that it will be possible to define a measure on G.

A transformation group is a group of mappings of a set X onto itself. We say that G has a left action on X (X is sometimes called a G-space) and write

$$\cdot : G \times X \longrightarrow X$$

by $g \cdot x = x'$ or simply $gx = x'$. G is said to act transitively on X (and X is said to be homogeneous) if for every $x_1, x_2 \in X$ there exists a $g \in G$ such that

$$x_2 = g x_1 .$$

Let X be a G-space (not necessarily transitive), then the set $\mathcal{O}_x = \{x' \mid x' = gx \; g \in G\}$ is called the orbit of x. Clearly

$\mathcal{O}_x$ is a homogeneous G-space and any set X can be partitioned into a pairwise disjoint union of orbits.

Fix an $x_0 \in X$ where X is a homogeneous G-space. The set $G_{x_0} = \{g \in G \mid gx_0 = x_0\}$ is obviously a subgroup. G_{x_0} is called the stabilizer of x_0. Let $G_{x_0} \cdot g^{-1}$ be a right coset of G_{x_0}. Then it is easy to see that gG_{x_0} is the set of all elements that take x_0 into $gx_0 = x_1$. Using this construction we can show that there is a bijection such that the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{\sim} & G/G_{x_0} \\ \downarrow & & \downarrow \\ X & \xrightarrow{\sim} & G/G_{x_0} \end{array}$$

where the vertical arrows are G-action on X and the natural G-action on G/G_{x_0} given by $g_1 : gG_{x_0} \longrightarrow g_1 gG_{x_0}$. If G is a topological group and X is a topological space we shall require

$$\cdot : G \times X \longrightarrow X$$

to be continuous under the product topology.

If in addition X is also a Borel space with measure μ then we shall require

$$g : x \longmapsto gx$$

to be measurable for all $g \in G$. The measure μ is said to be invariant under G-action if, if for any measurable set $B \subseteq X$,

$$\mu(B) = [g \cdot \mu](B) = \mu(gB).$$

An example is, of course, Lebesgue measure in R^n. It is

invariant under shifts

$$B \longmapsto x+B \qquad x \in R^n, \text{ measurable } B \subseteq R .$$

Invariance is a strict property. Often we have examples of G-actions on Borel spaces X such that X admits no nontrivial invariant measures. A relaxed condition is quasi-invariance, that is, if μ is equivalent to its translate $g \cdot \mu$

$$\mu \equiv g \cdot \mu$$

then μ is quasi-invariant. This condition says in other words that null sets are preserved under G-action.

If G is a locally compact topological group, then it is possible to find a left-invariant measure on G (under the natural left action of G on itself) (Loomis 1953, Halmos 1950).

Theorem (Haar). If G is a locally compact topological group then there exists a unique (up to multiplicative constant) left-invariant measure μ on G.

The theorem works of course for right invariance too. The measure μ is called left (or right) Haar measure.

If μ is left Haar measure then the measure defined as

$$\mu \cdot g(B) = \mu(Bg) \qquad \text{measurable } B \subseteq G,\ g \in G$$

(i.e. the right translate) is also left-invariant, hence by Haar's theorem there exists a constant such that

$$\Delta(g)\mu = \mu \cdot g.$$

This constant Δ is called the modular function of G.

2.3 Examples

We will discuss six groups that are of interest to vision.

Example 1. VG_0 is the group of translations on R^2 $x \longmapsto x+\alpha$.

Example 2. We will write VG for the group given by matrices

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \qquad a>0,\ b\in R$$

This group is commonly called the ax+b group or the group of homogeneous affinities of the line. Let $\begin{pmatrix} x \\ w \end{pmatrix}$ be homogeneous coordinates for the projective line. Then operating on the column vector by the matrix induces the action

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix} = \begin{pmatrix} ax+bw \\ w \end{pmatrix}$$

or letting $x' = \frac{x}{w}$

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ 1 \end{pmatrix} = \begin{pmatrix} ax'+b \\ 1 \end{pmatrix}.$$

Example 3. Let VG_2 be given by the matrices

$$\begin{pmatrix} a & 0 & b_1 \\ 0 & a & b_2 \\ 0 & 0 & 1 \end{pmatrix} \qquad a>0,\ b_1, b_2 \in R.$$

Letting $\begin{pmatrix} x_1 \\ x_2 \\ w \end{pmatrix}$ be homogeneous coordinates of the plane, it is easy to see that VG_2 is the group consisting of translations and dilatations of the plane.

Example 4. Let VG_3 be given by the matrices

$$\begin{pmatrix} \cos\theta & \sin\theta & b_1 \\ -\sin\theta & \cos\theta & b_2 \\ 0 & 0 & 1 \end{pmatrix} \qquad b_1, b_2 \in R,\ 0 \leq \theta < 2\pi.$$

This is the group of rotations and translations in the plane.

Example 5. Let VG_4 be the group of translations, dilatations and rotations in the plane:

$$\begin{pmatrix} a\cos\theta & a\sin\theta & b_1 \\ -a\sin\theta & a\cos\theta & b_2 \\ 0 & 0 & 1 \end{pmatrix} \qquad a>0,\ 0 \leq \theta < 2\pi,\ b_1, b_2 \in R.$$

Example 6. Finally, let VG_5 be the gamma group, i.e. the group that performs on an image $f(x,y)$ the folowing operation:

$$f(x,y) \longmapsto f(x,y)^{\gamma} \qquad \gamma > 0.$$

2.4 Representations

The mathematical expression of symmetry is given through the concept of group representations. More precise definitions follow, but we will first discuss the intuitive significance of these concepts.

We will be interested in some set of operations on the space of images, for example, the set of all translations of an image. Now it is not only important for us to know that this set of transformations forms a group, but also crucial to understand precisely which of

the many possible abstract group structures this set possesses. The concept of a group representation does this: It distills the abstract group structure from the unnecessary details bound up with the set of transformations and supplies those missing details by giving a map from the abstract group into the group of all linear transformations on a given space. Thus to each group element $g \in G$ the representation π takes g into a linear operator $\pi(g)$. For example, the set of translations possesses the abstract group structure of $(R,+)$, the real numbers under the operation of addition. The representation π in this case takes a single real number x and maps it into $\pi(x)$, the linear operator that translates an image by x units to the left.

The object of the symmetry theory is to make statements about the structure of certain operators (like the vision operator on the space of images). That is, we wish to examine extensively those operators satisfying a symmetry condition under our given set of transformations. To give meaning to the concept of symmetry in the language of group representations we use the definition of intertwining operator: The situation starts with two (perhaps non-distinct) representations π,π' along with an operator A that maps between their respective spaces (the spaces that the operators of the representation act upon). An operator is an intertwining operator if $\pi' A = A\pi$, i.e.

if a symmetry operation $\pi(g)$ is performed on a point v of the first space and then mapped by A onto the other space, then the result is identical to the mapping the point v by A into the second space and performing the corressponding symmetry operation $\pi'(g)$. In terms of our example, the first and second spaces are identical as are the representations (in this case an intertwining operator is also called a commuting operator), both representations being translation by an amount x. The condition of intertwining can then be easily seen to collapse to the definition of shift-invariance. Thus when speaking of intertwining one can consider it to be a general invariance condition.

Definition. A unitary representation of a locally compact topological group G on a Hilbert space H is a homomorphism of G into the group of automorphisms of H

$$\pi : G \longrightarrow GL(H)$$

with the property that, for each $x \in H$, the map of G into H given by

$$g \longmapsto \pi(g)x$$

is strongly continuous.

All Hilbert spaces will be assumed to be separable.

Examples: 1) If H is a finite dimensional space (say n-dimensional), then $GL(H) \cong GL_n$, the group of unitary nxn matrices under matrix multiplication, and π corresponds to the classical idea of matrix representations of groups.

2) A particularly important representation for us will be the following:

Let X be a locally compact G-space with invariant measure μ. Let $H=L^2(X,\mu)$ be the space of real- (or complex-) valued functions defined on X that are L^2 integrable with respect to the measure μ. Then define π by

$$\pi(g):f(x)\longmapsto f(gx).$$

It is easily seen that $\pi(g)$ is an invertible linear operator; furthermore, with an invariant measure, π is unitary.

An application of the latter example can be seen by letting X be the ordinary Euclidean plane R^2 and G be the group VG_3. We have that the numbers (θ,a,b) parameterize the group G. The space H is the Hilbert space of real-valued square integrable functions $L^2(R^2,\mu)$ which we take as a model for images. The representation

$$\pi(\theta,b_1,b_2)f(x_1,x_2)=f(x_1\cos\theta+x_2\sin\theta+b_1,-x_1\sin\theta+x_2\cos\theta+b_2)$$

is unitary when μ is ordinary Lebesgue measure. This representation is a model for the geometric rotation and translation of images.

Another application of this form is an example which we will study extensively since it has characteristic qualities useful in studying vision. Let X be the real line R and G be VG_1, the group of homogeneous affinities of the line; let $x\longmapsto ax+b$ where $a>0$ and a and b parameterize the group which we call the ax+b group. H is

the Hilbert space of real-valued square integrable functions $L^2(R,\mu)$. The representation $\pi(g)f(x)=f(ax+b)$ is not unitary under any measure (although it can be made to be unitary with slight modifications).

The reason we study representations is to study the operators on the Hilbert space H that satisfy a certain symmetry condition.

Definition. If π,π' are two representations of a group G on Hilbert spaces H, H' respectively, an intertwining operator A from H to H' is a bounded linear operator such that $\pi'(g)A=A\pi(g)$ for all $g\in G$. The set of all intertwining operators obviously forms a vector space -- we will write it as $\mathcal{R}(\pi,\pi')$. It turns out that $\mathcal{R}(\pi,\pi)$ is a weakly closed subalgebra of the Banach algebra of all bounded linear operators on H, which is called the commuting (von Neumann) algebra of π on H.

Definition. Two representations π,π' on H,H' respectively are equivalent if there is an isomorphism $A:H\longrightarrow H'$ intertwining π and π'. If A is unitary then π and π' are unitarily equivalent. The dimension of $\mathcal{R}(\pi,\pi')$ as a vector space is called the intertwining number of π and π'. When the intertwining number is 0 then π and π' are said to be disjoint.

If $H_1\subseteq H$ is a closed subspace and $\pi(g)x\in H_1$ for all $x\in H_1$, then we say that H_1 is an invariant subspace of H. Evidently $\pi(g)$ restricted to H_1 for all $g\in G$ gives another

representation π_1 of G which is called a subrepresentation of π. It can happen that π has no proper subrepresentations (trivial subrepresentations are π itself and the representation taking every g to the zero operator). If this is the case then π is irreducible. If π admits an invariant subspace H_1 such that H_1 admits an invariant complement H_2, then we say that π is completely reducible and we write $\pi = \pi_1 \oplus \pi_2$ where π_1 and π_2 are the restrictions to H_1 and H_2. This is called a direct sum decomposition of π.

The contact point between group representation theory and the desired structure theory of intertwining operators is the very important

Schur's Lemma. Let π_1, π_2 be two irreducible representations of G and let A be an intertwining operator; then A is either an isomorphism or 0.

Proof. Ker A, Im A are invariant subspaces. If π_1, π_2 are irreducible then they cannot have any proper subrepresentations. Hence Ker A, Im A are either {0} or the entire space.

Corollary. Two irreducible representations are either equivalent or disjoint.

This lemma is crucial to the structure theory because it severely limits the structure of operators intertwining or commuting with irreducible representations. The key idea behind the structure theory is to decompose a given

pair of representations into a sum of irreducibles. Loosely speaking, the intertwining operator may then be considered to be a collection of maps from the set of irreducible subspaces in the domain to its counterpart in the range. Schur's lemma assures us that the elements of this collection of maps is of a simple structure.

CHAPTER 3

LINEAR SYMMETRY THEORY

3.1 Introduction

The spectral theorem in elementary functional analysis helps illustrate what is meant here by a structure theory. For a normal operator A the spectral theorem finds a projection-valued measure E defined on the spectrum σ of A such that

$$A=\int_{\sigma} \zeta \, dE(\zeta) .$$

This is the celebrated spectral resolution of a normal operator. Much of the usefulness of this formula stems from the fact that the commutative C^*-algebra generated by A is given by a functional calculus on the spectrum σ of A. For example,

$$\exp(A)=\int_{\sigma} e^{\zeta} dE(\zeta) .$$

Note that although the integrand in the spectral resolution varies with the operator, the projection-valued measure is fixed. This projection-valued measure provides the "structure" of the algebra while the integrand specifies a given element.

This chapter will describe methods for developing a structure for an unknown operator A. That is, A will be expressed as

$$A=\int_{\Lambda} a(\lambda)\,dE(\lambda)\,.$$

Symmetry considerations allow us to deduce the projection-valued measure, the "structure" of the operator, without knowing what the operator is. Thus system identification is reduced to measuring the function $a(\lambda)$ -- this must be left to the experimenter. We will be working toward a proof of the

Fundamental Structure Theorem. Let π be a unitary representation on a Hilbert space $\mathcal{H}$. Let $A\in\mathcal{R}(\pi,\pi)$ be a normal operator. Then there is a spectral resolution of A

$$A=\int_{\Lambda} a(\lambda)\,dE(\lambda)$$

corresponding to a direct integral decomposition

$$\mathcal{H}=\int_{\Lambda}^{\oplus}\mathcal{H}_{\lambda}\,d\mu(\lambda)$$

where

$$\pi=\int_{\Lambda}^{\oplus}\pi_{\lambda}\,d\mu(\lambda)$$

and the π_{λ} are μ-a.e. irreducible.

The bulk of the material in this chapter will explain the meaning of the terms in this theorem and develop a method for constructing such a decomposition. This theorem is new and will be directly applicable to our formal modelling efforts. To say this theorem is new is not to say, however, that it represents an advance in the mathematical art. Rather, we emphasize that this theorem collects some well-known facts in a different light to facilitate the aims of formal modelling.

3.2 Finite-dimensional Case

In this section we shall show how the structure theory works for the case of a finite dimensional representation with a commuting operator which is a normal linear transformation, i.e. a matrix that commutes with its conjugate transpose. First, we shall decompose the vector space into a finite number of bricks (irreducible subspaces). Second, we show how every commuting operator must be diagonalized by these bricks. This then is the structure theory -- the basis vectors contained in a single brick diagonalize the matrix with the same value along the diagonal. This follows from the symmetry. These basis vectors are of course, eigenvectors of the operator and therefore give the structure of the commuting operator. Their corresponding eigenvalues, i.e. the numbers along the diagonal must be measured by experiment.

Given a finite-dimensional unitary representation on a space V, how can we decompose it (in the sense of Section 2.3) into irreducible subspaces? Well, if there exist no proper invariant subspaces then V is itself irreducible. Otherwise we can find an invariant subspace V' of V. Taking the orthogonal complement $V'^{\perp}$, it is easy to see that $V' \oplus V'^{\perp}$ decomposes π. Repeat this procedure for V' and $V'^{\perp}$. Because V is finite-dimensional there cannot exist a nonterminating sequence of proper subspaces $V \subsetneq V' \subsetneq V'' \subsetneq V''' \subsetneq \cdots$. Thus there exists a finite set $\{V_i\}_{i \in I}$

(I is a finite index set) of irreducible invariant subspaces such that $\bigoplus_{i\in I} V_i = V$ and $V_{i_1} \cap V_{i_2} = 0$. Consider the set of projections $\{E_i\}_{i\in I}$ on each of the invariant irreducible subspaces. We have the properties $\sum_{i\in I} E_i = 1_V$ (the identity transformation) and $E_i E_j = E_j E_i = 0$ if $i \neq j$. Schur's lemma gives us that each projection E_i is minimal. That is, if E is another commuting projection such that $E \leq E_i$, i.e. $\mathrm{im}\{E\} \subseteq \mathrm{im}\{E_i\}$, then $E = E_i$ or $E = 0$.

Now consider the spectral resolution for an unknown commuting operator $A = \sum_{\alpha\in J} \lambda_\alpha \tilde{E}_\alpha$ where J indexes the eigenvalues λ_α of A and $\tilde{E}'_\alpha$ is the projection onto the eigenspace of λ_α. For any $\tilde{E}'_\alpha$ we have

$$\tilde{E}'_\alpha = \tilde{E}'_\alpha(1_V) = \tilde{E}'_\alpha(E_1 + E_2 + E_3 + \ldots + E_n) = \tilde{E}'_\alpha E_1 + \tilde{E}'_\alpha E_2 + \tilde{E}'_\alpha E_3 + \ldots + \tilde{E}'_\alpha E_n.$$

But for each term we have

$$\tilde{E}'_\alpha E_i \leq E_i .$$

Thus $\tilde{E}'_\alpha E_i = 0$ or $\tilde{E}'_\alpha E_i = E_i$. Let I_α denote all indices i where the latter case occurs. Then

$$\tilde{E}'_\alpha = \sum_{i\in I_\alpha} E_i .$$

It is obvious that $I_\alpha \cap I_{\alpha'} = 0$ if $\alpha \neq \alpha'$ and $\bigcup_{\alpha\in J} I_\alpha = I$. Hence for every α there corresponds a unique i and every i arises from a unique α, so that

$$A = \sum_{\alpha\in J} \lambda_\alpha \tilde{E}'_\alpha = \sum_{\alpha\in J} \lambda_\alpha \sum_{i\in I_\alpha} E_i = \sum_{i\in I} \lambda'_i E$$

where $\lambda'_i = \lambda_\alpha$ when $i \in I_\alpha$.

This gives us a structure for the unknown A expressed by the projection-valued measure $\{E_i\}_{i\in I}$. To actually pin down the operator A we must measure the eigenvalues λ'_i by

experiment.

3.3 Infinite-dimensional Case

As may be supposed, the situation in the infinite-dimensional case is a good deal more delicate than the finite-dimensional case -- even when considering separable Hilbert spaces (with which we shall deal exclusively). The following example displays some of the difficulties.

Consider the representation π of the additive group R where V is $L^2(R,\mu)$ (μ Lebesgue). We let

$$\pi(x): f(y) \longmapsto f(y+x) \qquad y \in R.$$

Taking the Fourier transform we have the unitarily equivalent representation ($\hat{}$ is unitary)

$$\hat{\pi}(x): \hat{f}(\omega) \longmapsto \hat{f}(\omega)\, e^{i\omega x} \qquad \omega \in R.$$

Let B be any measurable subset of R, then the characteristic function of B, I_B, induces a projection onto an invariant subspace given by E(B):

$$\hat{f}(\omega) \xmapsto{E(B)} I_B(\omega) \cdot \hat{f}(\omega) \qquad \omega \in R.$$

It is not hard to show that every commuting projection arises in this manner. Let's examine the subrepresentation given by B. Now if B has positive measure then there is a set $B_1 \subseteq B$ such that $\mu(B_1) < \mu(B)$. Therefore $E(B_1)$ projects onto a proper subspace so that the representation $\hat{\pi}\big|_{E(B)}$ is reducible. If B has measure zero then the range of E(B) is zero.

Thus we have the result that $\hat{\pi}$ (and π), which are perfectly natural and interesting representations, have no irreducible representations! Our representation thus has no direct sum decomposition into irreducible representations. This state of affairs explains why the fundamental structure theorem is stated in terms of direct integrals instead of a direct sum (note that even a countable direct sum will not do).

3.4 Direct Integrals

We want to define an analog of direct sum for Hilbert spaces -- a "continuous" direct sum. We start with a locally compact measure space (Λ, μ) and a collection of Hilbert spaces $\{H_\lambda\}_{\lambda \in \Lambda}$.

First we define a direct integral for the special case H_λ equal to some fixed space H_0 for almost all λ.

Definition. A vector-valued function $\xi : \Lambda \longrightarrow H_0$ is a measurable vector field if, for any $v \in H_0$, $\langle \xi(\lambda), v \rangle$ is a measurable complex-valued function on Λ.

Proposition. If $\xi(\lambda)$, $\eta(\lambda)$ are measurable vector fields, then $\langle \xi(\lambda), \eta(\lambda) \rangle$, $\|\xi(\lambda)\|$ are measurable functions.

Proof. Since H_0 is separable, we can find a countable basis $\{v_k\} \subseteq H_0$. So we have

$$\xi(\lambda) = \sum_k \langle \xi(\lambda), v_k \rangle v_k$$

and

$$\langle \xi(\lambda), \eta(\lambda) \rangle = \sum_k \langle \xi(\lambda), v_k \rangle \langle v_k, \eta(\lambda) \rangle .$$

The second assertion follows.

Definition. The direct integral $H=\int_\Lambda^\oplus H_\lambda d\mu(\lambda)$ is the Hilbert space H of all measurable vector fields $\xi(\lambda)$ such that

$$\left| \int_\Lambda \|\xi(\lambda)\|^2 d\mu(\lambda) \right| < \infty.$$

It is easy to see that (with the usual identification of functions differing on null sets) $\int_\Lambda^\oplus H_\lambda d\mu(\lambda)$ is a Hilbert space, i.e. H is complete, etc.

We introduce a coordinate notation for the direct integral with respect to a countable orthonormal basis $\{v_k\}$ of H_0.

Let $a_k(\lambda)=\langle \xi(\lambda), v_k \rangle$. Then to each ξ corresponds this sequence of measurable functions.

Proposition. $\xi(\lambda) \in \int_\Lambda^\oplus H_\lambda d\mu(\lambda)$ iff $\sum_k \int_\Lambda |a_k(\lambda)|^2 d\mu(\lambda) < \infty$.

Proposition. For $\xi(\lambda), \eta(\lambda) \in \int_\Lambda^\oplus H_\lambda d\mu(\lambda)$, we have $\langle \xi(\lambda), \eta(\lambda) \rangle = \sum_k \int_\Lambda a_k(\lambda) \overline{b_k(\lambda)} d\mu(\lambda)$ where $a_k(\lambda)=\langle \xi(\lambda), v_k \rangle$ and $b_k(\lambda)=\langle \eta(\lambda), v_k \rangle$.

To generalize the above definition to direct integrals involving Hilbert spaces H_λ where H_λ can vary in dimension, we first fix spaces $\tilde{H}_1 \subseteq \tilde{H}_2 \subseteq \tilde{H}_3 \subseteq \ldots \subseteq \tilde{H}_\infty$ (for example let $\tilde{H}_\infty = l^2$, the set of sequences $(x_1, x_2, \ldots)$ such that $\sum |x_k|^2 < \infty$, and $\tilde{H}_m$ is the l^2 subspace consisting of sequences of nonzero elements only in the first m coordinates). Next let $\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_\infty$ (called dimension

sets) be a collection of disjoint Borel sets of a locally compact measure space (Λ,μ) fixed for the remainder of this discussion. A vector-valued function $f(\lambda)$ with $f(\lambda)\in\tilde{H}_m$ if $\lambda\in\tilde{e}_m$ will be called a measurable vector field if, f restricted to $\tilde{e}_n$, $f|_{\tilde{e}_n}$ is a measurable vector field in the old sense for each n.

Definition. $\int_\Lambda^\oplus H_\lambda d\mu(\lambda)$ is the set of measurable vector fields such that

$$\left|\sum_n \int_{\tilde{e}_n} \|\xi(\lambda)\|^2_{\tilde{H}_n} d\mu(\lambda)\right| < \infty.$$

The inner product of $\xi(\lambda),\eta(\lambda)\in\int_\Lambda^\oplus H_\lambda d\mu(\lambda)$ is defined by $\langle\xi(\lambda),\eta(\lambda)\rangle = \sum_n \int_{\tilde{e}_n}\langle\xi(\lambda),\eta(\lambda)\rangle d\mu(\lambda)$.

A coordinate representation may be defined with respect to a countable orthonormal basis $\{\tilde{v}_k\}$ of $\tilde{H}_\infty$ with the condition that $\{\tilde{v}_1,\ldots,\tilde{v}_m\}$ is a basis of $\tilde{H}_m$ for all m.

Let $a_k(\lambda)=\begin{cases}\langle\xi(\lambda),v_k\rangle & \text{if } \lambda\in e_m \text{ and } k\le m\\ 0 & \text{otherwise.}\end{cases}$

Then $\int_\Lambda^\oplus H_\lambda d\mu(\lambda)$ is isomorphic to a Hilbert space of a sequence of measurable functions

$$(a_1(\lambda),a_2(\lambda),\ldots)$$

satisfying

i) $\left|\sum_k \int_\Lambda |a_k(\lambda)|^2 d\mu(\lambda)\right| < \infty$

ii) $a_k(\lambda)=0$ for $\lambda\in\tilde{e}_m$, $k>m$.

An operator A on $H=\int_\Lambda^\oplus H_\lambda d\mu(\lambda)$ is said to be <u>decomposable</u> if there exists a family of operators $\{A_\lambda\}$ where A_λ acts on H_λ such that

$$A[f(\lambda)] = A(\lambda)[f(\lambda)] \qquad \text{a.e. } \Lambda$$

A decomposable operator A is <u>diagonal</u> if for almost all we have $A(\lambda) = a(\lambda) \cdot 1_{H_\lambda}$, i.e. $A(\lambda)$ is a scalar multiple of the identity.

3.5 Direct Integral Decompositions

The representations π_λ on H_λ are said to be a direct integral decomposition of a representation π on H, and we write

$$\pi = \int_\Lambda^\oplus \pi_\lambda \, d\mu(\lambda)$$

if

a) $H \simeq \int_\Lambda^\oplus H_\lambda \, d\mu(\lambda)$

b) the operators $\pi(g)$, $g \in G$, are decomposable,

i.e. $\pi(g)[f(\lambda)] = \pi_\lambda(g)[f(\lambda)]$.

Going back to the original example, it is clear that the representation $\hat{\pi}$ on $L^2(R,\mu)$ is decomposed as a direct integral of the one-dimensional representations $\hat{\pi}_\omega$ on $\hat{H}_\omega \simeq \mathbb{C}$ given by

$$\hat{\pi}_\omega(x)\xi = e^{i\omega x} \cdot \xi \qquad \text{where } \xi \in H_\omega = \mathbb{C}.$$

How can we tell if a given representation has a direct integral decomposition? Remember that there exists a certain duality between a finite direct sum decomposition of a space $V = \bigoplus_{i=1}^n V_i$ and the set of projections $\{E_i \mid E_i \text{ is onto } V_i\}$. This observation is the key to direct integral representations. In fact, the set of projections $\{E_i\}$ and its correspondence with the direct

sum decomposition $V=\bigoplus_{i=1}^{n} V_i$ can be generalized to a correspondence between a Boolean algebra of projections and a direct integral decomposition. This motivates the following

Definition. A commutative set Q of projections in a Hilbert space V is a Boolean algebra of projections if it contains the zero projection, the identity projection 1, and contains along with any E_1 and E_2 the projections $1-E_1$ and E_1E_2. A Boolean algebra of projections is complete if, along with any subfamily $\{E_\alpha\}$ of Q, Q contains the least upper bound of $\{E_\alpha\}$.

A canonical example of a complete Boolean algebra is the following familiar object from functional analysis.

Definition. Let X be a Borel space and let E be a function which carries each Borel set B in X into a projection E(B) in a Hilbert space H such that

1) $E(B_1)E(B_2)=E(B_1\cap B_2)$

2) if $B_i\cap B_j=\phi$ for $i\neq j$, then $\sum_j E(B_j)=E(\bigcup_j B_j)$

3) $E(\phi)=0$, $E(X)=1$.

Then E is called a projection-valued measure on X to H.

Lemma. The range of a projection-valued measure is a complete Boolean algebra of projections.

We now will establish the duality between a direct integral decomposition of a representation and a complete Boolean algebra of projections.

Given $\pi = \int_X^{\oplus} \pi_x d\mu(x)$ and $H = \int_X^{\oplus} H_x d\mu(x)$, for each Borel set $B \subseteq X$ let $E(B)$ be the projection defined by

$$f(x) \xmapsto{E(B)} I_B(x) \cdot f(x)$$

where I_B is the characteristic function of the set B. Note that the range of E is a Boolean algebra of projections Q. There are two crucial properties of Q: a) $Q \subseteq \mathcal{R}(\pi, \pi)$, b) Q is a complete Boolean algebra. To exhibit the other (hard) side of the duality, we now start with a representation that admits a Boolean algebra B satisfying conditions a) and b) and end up with a direct integral decomposition of the representation π.

3.6 The Group Algebra

To every locally compact group G we can associate a Banach algebra of a certain type, called C^*-algebras. This association will allow us to use the techniques of C^*-algebras in obtaining a structure theory.

Definition. A C^*-algebra is a Banach algebra with involution, with the following extra condition:

(*) $\|x^*x\| = \|x\|^2$ for every element x of the C^*-algebra.

The only C^*-algebra in which we will be interested is associated with a locally compact group G. Let μ be (left) Haar measure on G. Then $L^1(G)$, the space of absolutely integrable complex-valued functions, forms a Banach algebra under the multiplicative operation of

convolution:

$$f*g(s)=\int_G f(t)\,g(t^{-1}s)\,d\mu(t) \qquad f,g\in L^1(G) \quad s,t\in G.$$

It is easy to check that $L^1(G)$ is a Banach algebra under the usual norm,

$$\|f\|=\int_G |f(t)|\,d\mu(t) \qquad f\in L^1(G) \quad t\in G.$$

An involution is defined as

$$f^*(t)=\Delta(t^{-1})\,\overline{f(t^{-1})},$$

where Δ is the modular function defined in Section 2.2. It is easy to see that

$$\|f\| = \|f^*\|$$

but that $L^1(G)$ does not form a C^*-algebra because $(*)$ does not hold. To make $L^1(G)$ into a C^*-algebra, we define a new norm

$$\|f\|_1=\sup_{\pi}\|\pi(f)\|$$

where the supremum ranges over all representations of $L^1(G)$ that preserve the involution. Then let $C^*(G)$ be the completion of $L^1(G)$ under $\|f\|_1$.

What relation exists between a representation of a given group G and a representation of its group algebra?

Lemma (Loomis 1953). If π is a unitary representation of G on a Hilbert space then we can define a corresponding unitary representation of $L^1(G)$ that preserves involution.

Proof. Let $\pi(f)=\int_G f(s)\pi(s)\,ds$.

Lemma (Loomis 1953). Conversely if π is a representation of $L^1(G)$ on H, preserving involution such that the union

of the ranges of the operators $\pi(f)$, $f \in L^1(G)$ is dense in H, then we can define a unitary representation π on G.

Proof. Let $u_1, u_2, \ldots$ be an approximate identity in $L^1(G)$; then define $\pi(g) = \lim_{n\to\infty} \pi(u_n^g)$ where $u_n^g(s) = u_n(gs)$, i.e. the left translate of u_n.

These constructions can of course be extended to the C^*-algebra $C^*(G)$. Unless G is discrete, there exist only approximate identities but no genuine identity e of the C^*-algebra. This is easily rectified by formally adjoining an identity and extending a representation π by

$$\pi(\gamma e + f) = \gamma 1_H + \pi(f).$$

3.7 Positive Linear Functionals of a C^*-algebra

A functional f of a C^*-algebra with unit e is a positive linear functional if $f(xx^*) \geq 0$. There is a well-known correspondence between positive linear functionals and representations that we will need.

Given a representation π of a C^*-algebra on a Hilbert space H, we see that $f(x) = \langle \pi(x)\xi, \xi \rangle$ is positive. Indeed,

$$\langle \pi(x^*x)\xi, \xi \rangle = \langle \pi(x^*)\pi(x)\xi, \xi \rangle = \langle \pi(x)^*\pi(x)\xi, \xi \rangle$$

$$= \langle \pi(x)\xi, \pi(x)\xi \rangle = \|\pi(x)\xi\| \geq 0.$$

Conversely, given a positive linear functional f, we can construct a representation by using a procedure due to Gelfand, Naimark and Segal (the GNS construction). Roughly the procedure is as follows: 1) Take the left

regular representation $\pi_0: A \longrightarrow A$, i.e. $\pi_0(x)y = xy$. Clearly π_0 is a representation. 2) We define a sequilinear form on A by $\langle x,y\rangle = f(y^*x)$. Clearly $\langle \pi_0(x)y, z\rangle = \langle y, \pi_0(x^*)z\rangle$. 3) To make $\langle \cdot,\cdot \rangle$ an inner product, we factor out $J = \{x \mid f(x^*x) = 0\}$.

J turns out to be an ideal so we can lift π_0 to a representation π of A/J.

To see that π is bounded, $H \simeq A/J$, see Arveson (1976). Note that $e+J \in A/J = H$ is a cyclic vector and that there exists an embedding of A onto H given by the projection

$$x \longmapsto x+J.$$

If two cyclic representations π, ξ_0 and π', ξ_0' have identical positive linear functionals, then define $U\xi = \xi'$ where $\xi = \pi(x)\xi_0$, $\xi' = \pi'(x)\xi_0'$. U is unitary since

$$\langle U\xi_1, U\xi_2\rangle = \langle \xi_1', \xi_2'\rangle = \langle \pi'(x_1)\xi_0', \pi'(x_2)\xi_0'\rangle$$
$$= \langle \pi'(x_2 x_1^*)\xi_0', \xi_0'\rangle = f(x_2^* x_1)$$

and

$$\langle \xi_1, \xi_2\rangle = \langle \pi(x_1)\xi_0, \pi(x_2)\xi_0\rangle = \langle \pi(x_2^* x_1)\xi_0, \xi_0\rangle = f(x_2^* x_1).$$

3.8 The Spectral Decomposition of a Representation of a Commutative C^*-algebra (Cyclic Case)

From the Gelfand-Naimark theorem in functional analysis (Rudin 1973, p.276) recall that the Gelfand transform $x \longmapsto \hat{x}$ for $x \in A$ is an isometric isomorphism of a commutative C^*-algebra A onto the space of complex-valued

continuous functions $C(\Delta)$ defined on Δ, the maximal ideal space of A. We follow a treatment given by Godement and Mackey (Maurin 1968). Recall that π is called a cyclic representation if there exists a vector $\xi_0 \in H$ such that

$$\pi(A)\xi_0$$

is dense in H.

There is an intertwining isomorphism U of a cyclic representation (π, ξ_0) of a commutative C^*-algebra A on H to a subspace L of $L^2(\Delta,\mu)$.

$$U: H \longrightarrow L^2(\Delta,\mu)$$

The representation π' is given by

$$\pi'(x):[\varphi(\lambda)] \longmapsto \hat{x}(\lambda)\cdot\varphi(\lambda)$$

where $x \in A, \lambda \in \Delta$, $\varphi(\cdot) \in L^2(\Delta,\mu)$, where $x \longmapsto \hat{x}$ is the Gelfand isomorphism of A onto $C(\Delta)$, the space of all continuous complex-valued functions defined on Δ.

We know $f(x)=\langle \pi(x)\xi_0, \xi_0\rangle$ is a positive linear functional. Thus for $x \geq 0$ (i.e. $\hat{x}(\lambda) \geq 0$) we can find an $\hat{x}^{\frac{1}{2}}(\lambda) \ni \hat{x}^{\frac{1}{2}}(\lambda)\,\hat{x}^{\frac{1}{2}}(\lambda) = \hat{x}(\lambda)$. Let $x^{\frac{1}{2}} \in A$ be the element that the Gelfand transform takes into $\hat{x}^{\frac{1}{2}}(\lambda)$. Then $x^{\frac{1}{2}}x^{\frac{1}{2}} = x$, $f(x) = \langle \pi(x^{\frac{1}{2}*}x^{\frac{1}{2}})\xi_0, \xi_0\rangle = \|\pi(x^{\frac{1}{2}})\xi_0\|^2 > 0$.

Thus for $\hat{x}(\lambda) \geq 0$, we can define a positive linear functional $\hat{f}(\hat{x}(\lambda)) = f(x) \geq 0$. This positive linear functional can be extended to real- and complex-valued functions in the usual way. Invoking the Riesz representation theorem we find a measure μ such that

$$\hat{f}(\hat{x}(\lambda)) = \int_{\Delta} \hat{x}(\lambda)\, d\mu(\lambda).$$

Let x_ξ, x_η be such that $\xi=\pi(x_\xi)\xi_0$, $\eta=\pi(x_\eta)\xi_0$. Define the isomorphism U by

$$U:\xi \longmapsto \hat{x}_\xi(\lambda).$$

Since π is faithful, U is well defined and one-to-one.

$$\langle U\xi, U\eta\rangle=\langle \hat{x}_\xi(\lambda), \hat{x}_\eta(\lambda)\rangle=\int_\Delta \hat{x}_\xi(\lambda)\overline{\hat{x}_\eta(\lambda)}\,d\mu$$

$$=f(\hat{x}_\xi(\lambda)\overline{\hat{x}_\eta(\lambda)})=f(x_\xi x_\eta^*)=\langle\pi(x_\xi x_\eta^*)\xi_0,\xi_0\rangle$$

$$=\langle\pi(x_\eta^* x_\xi)\xi_0,\xi_0\rangle=\langle\pi(x_\xi)\xi_0,\pi(x_\eta)\xi_0\rangle=\langle\xi,\eta\rangle$$

so that U is an isometry onto a subspace L of $L^2(\ ,\ )$.

3.9 Direct Integral Decompositions of a Commutative C*-algebra

If π is a representation of A on a Hilbert space H with countable basis $\{v_k\}$, then we can decompose π into the direct sum of cyclic representations.

$$H=\bigoplus_j H_j \qquad \pi=\bigoplus_j(\pi_j,\xi_j)$$

Indeed, let H_1=closure of $\pi(A)v_1$. If $H_1=H$ then H is cyclic and we are done, else let E_2 be the projection onto the orthocomplement of H_1 in H, i.e. $H\ominus H_1$. Let $H_2=\pi(A)E_2v_2$, and so on.

By the section above we can find for each H_k an isometric isomorphism U_k, a subspace of $L^2_{\mu_k}(\Delta)$. Thus there exists an isometric isomorphism U of H onto $\oplus L^2_{\mu_k}(\Delta)$. Moreover the operators $\pi(x)$ map under to U to operators $\tilde{x}$ of pointwise multiplication by $\hat{x}(\lambda)$. For each k we have a positive functional $f_k(x)=\langle\pi_k(x)\xi_k,\xi_k\rangle$. Define a new functional by

(3.9.1) $$f(x)=\sum_{k=1}^{\infty}\frac{1}{2^{k+1}}f_k(x).$$

This is again a positive functional on functions of Δ so that we can define by the Riesz representation theorem a positive measure μ such that

$$f(x)=\int_{\Delta}\hat{x}(\lambda)\,d\mu.$$

If a set $M\subseteq\Delta$ is of μ-measure zero, i.e. if

$$\mu(M)=\int_{\Delta}I_M(\lambda)\,d\mu=0$$

then it follows that $f(I_M(\lambda))=0$, so that each term in 3.9.1) is zero, i.e. $\mu_k(M)=0$. Thus we have that $\mu_k\leq\mu$ for all k. Thus by the Radon-Nikodym theorem there exists a positive function $w_k=\frac{d\mu_k}{d\mu}$. Let $\tau_k(\lambda)=w_k^{\frac{1}{2}}(\lambda)$. Then we construct the isometric isomorphism U of H into the direct sum $H=\int_{\Lambda}^{\oplus}H_\lambda\,d\mu(\lambda)$ where $H_\lambda=l^2$.

First let $u=\sum_i\pi(x_i)\xi_i$, $v=\sum_k\pi(y_k)\xi_k$, where $x_i, y_k\in A$. We define a transformation U_λ for each $\lambda\in\Lambda$:

$$U: u\longmapsto(\tau_1(\lambda)\hat{x}_1(\lambda),\tau_2(\lambda)x_2(\lambda),\ldots)\in l^2.$$

We show for a fixed $u\in H$ that

$$U_\lambda u\in\int_{\Lambda}^{\oplus}H_\lambda d\mu(\lambda).$$

$$(u,v)=\sum_{i,k}\langle\pi(x_i)\xi_i,\pi(y_k)\xi_k\rangle=\sum_{i,k}\langle\pi(y_k)^*\pi(x_i)\xi_i,\xi_k\rangle$$

$$=\sum_i\langle\pi(y_i)^*\pi(x_i)\xi_i,\xi_i\rangle \text{ since } \xi_k\perp\pi(A)\xi_i$$

$$=\sum_i\int_{\Lambda}\hat{x}_i(\lambda)\overline{\hat{y}_i(\lambda)}\,d\mu_i(\lambda)=\sum_i\int_{\Lambda}\hat{x}_i(\lambda)\overline{\hat{y}_i(\lambda)}\,w_i(\lambda)\,d\mu$$

$$=\sum_i\int_{\Lambda}\hat{x}_i(\lambda)\overline{\hat{y}_i(\lambda)}\,\tau_i(\lambda)\,\tau_i(\lambda)\,d\mu$$

$$=\int_{\Lambda}\sum_i x_i(\lambda)\,\tau_i(\lambda)\,\overline{\hat{y}_i(\lambda)\,\tau_i(\lambda)}\,d=\int_{\Lambda}\langle U_\lambda u,U_\lambda v\rangle d\mu.$$

Thus U is isometric and it carries $H=\bigoplus_k \pi(A)\xi_k$ onto a subspace L of $\int_\Lambda^\oplus H_\lambda d\mu(\lambda)$. Let $F_\lambda = U_\lambda(H) \subseteq l^2$, then

$$L=\int_\Lambda^\oplus F_\lambda d\mu(\lambda).$$

We can find inductively a countable basis in L. Fix $\lambda \in \Delta$. Let $\eta_1(\lambda)=x_{v_1}(\lambda)$ where $\{v_k\}$ is the countable basis for the separable Hilbert space H. Given $\eta_1(\lambda),\ldots,\eta_k(\lambda)$ let $\eta_{k+1}(\lambda)$ be the first $\hat{x}_{v_i}(\lambda)$ linearly independent of $\eta_1(\lambda),\ldots,\eta_k(\lambda)$, if no such v_i exists let $\eta_{k+1}(\lambda)\equiv 0$. Since $\langle \hat{x}_{v_i}(\lambda), \hat{x}_{v_j}(\lambda)\rangle$ is measurable and the vanishing of the Gram determinant (which is also measurable)

$$G(\lambda)=\det\begin{pmatrix} \langle \hat{x}_{v_1}(\lambda), \hat{x}_{v_1}(\lambda)\rangle & \langle \hat{x}_{v_1}(\lambda), \hat{x}_{v_2}(\lambda)\rangle \cdots & \langle x_{v_1}(\lambda), \hat{x}_n(\lambda)\rangle \\ \langle \hat{x}_{v_2}(\lambda), \hat{x}_{v_1}(\lambda)\rangle & & \vdots \\ \vdots & & \vdots \\ \langle \hat{x}_{v_n}(\lambda), \hat{x}_{v_1}(\lambda)\rangle & & \langle \hat{x}_{v_n}(\lambda), \hat{x}_{v_n}(\lambda)\rangle \end{pmatrix}$$

determines the linear dependence of these vectors. The functions $\langle \eta_m(\lambda), \eta_n(\lambda)\rangle$ are measurable. Hence the function $d:\Lambda \longrightarrow Z^+ \cup \{\infty\}$ defined by $d(\lambda)=\dim F_\lambda = \sum_{n=1}^{\infty} \operatorname{sgn}\langle \eta_n(\lambda), \eta_n(\lambda)\rangle$ is measurable. Thus the dimension sets $\tilde{e}_n = d^{-1}(n)$ are Borel sets of Λ, $\eta_i(\lambda)$ are coordinate functions of a direct integral. Thus we have constructed a direct integral decomposition

$$U: H \longrightarrow \int_\Lambda^\oplus F_\lambda d\mu(\lambda)$$

such that the operators $\pi(x)$ are diagonal.

A structure theorem follows by the following observation:

Each Borel set B of Δ induces a projection-valued measure on $C(\Delta)$ through the characteristic function

$$\hat{f}(\lambda) \longmapsto I_B(\lambda) \cdot \hat{f}(\lambda).$$

Using the isomorphism U we can define a projection-valued measure $E(\cdot)$ on H, i.e.

$$E(B)\xi = U^{-1}\,[I_B(\lambda)\,\hat{x}_\xi(\lambda)]$$

where $\xi \in H$. Thus we have the structure formula

$$\pi(x) = \int_\Delta \hat{x}(\lambda)\, dE(\lambda) \qquad x \in A.$$

3.10 Commutative Groups

Let G be a locally compact Abelian group. If χ is a continuous homomorphism of G into T, the group of complex numbers of unit modulus with the operation of multiplication, then χ is called a <u>character</u>.

Notice that χ gives a one-dimensional unitary representation defined by

$$\pi(g)\,x = \chi(g) \cdot x \qquad x \in \mathbb{C}.$$

Also, because π is a representation on a one-dimensional space, π is irreducible. Schur's lemma gives immediately that the <u>only</u> irreducible representations of an Abelian G are one-dimensional and are given in this form. Thus the set of all characters is $\hat{G}$.

We can classify all unitary representations of locally compact Abelian groups with the following theorem (Loomis 1953).

<u>Stone's Theorem</u>. If π is a unitary representation of the

locally compact Abelian group G on a Hilbert space H, then there is a projection-valued measure E(B) on $\hat{G}$ such that

$$\pi(g) = \int_{\hat{G}} \overline{\chi(g)}\, dE(\chi) .$$

3.11 Relations Between Different Decompositions

We now know that, given a Boolean subalgebra Q of the commuting algebra $\mathcal{R}(\pi)$ of a representation π on a Hilbert space H, we can decompose H into a direct integral such that the commutative algebra A generated by Q consists of diagonal operators. Because $\pi(G)$ is in the commutant of $\pi(A)$, we have that π maps G into decomposable operators. Note that although a structure theorem for A is in hand, we still do not have one for the entire algebra $\mathcal{R}(\pi)$. The remaining sections are devoted to developing the structure formula for a normal operator in $\mathcal{R}(\pi)$ and giving a procedure that chooses a felicitous Q. Although the general problem of choosing Q remains open, the procedure for choosing Q for a given subclass of groups (which fortunately contains all the groups of interest to perception so far) was given by Mackey long ago.

A natural question that arises is how the choice of Q affects the direct integral decompositions. The main theorem we want here is Mautner's (1950).

Theorem. Q induces a maximal commutative subalgebra A of $\mathcal{R}(\pi)$ iff π_λ is irreducible on H μ-a.e.

This is all the decomposition theory we will need to study perception. There are many important further issues, such as uniqueness of decompositions. But we have omitted them because we do not need them. For the same reasons, we have left out multiplicity theory.

3.12 Induced Representations

Up until now we have hardly used any group theory at all. In order to carry out the next step in our program, that is, to obtain a suitable Boolean algebra to decompose our representations, it will be necessary to "look inside" the groups we are interested in. Specifically, we take advantage of the fact that 1) representations of certain subgroups have easy-to-calculate maximal Boolean algebras; 2) there is a procedure of "building up" representations of certain groups given representations of their subgroups; 3) along with the building up procedure is a theorem giving us a desired projection-valued measure whose range is a Boolean subalgebra of the commuting algebra of the built-up representation. Thus these steps are crucial in giving us our direct integral decomposition. Steps 2) and 3) are due to Mackey's postwar efforts. 2) is called induction (Mackey 1952,1953) and 3) is the imprimitivity theorem (Mackey 1949).

We shall use without proofs some results in Mackey

(1952). First note that the space of right cosets X=G/H of a subgroup H in a group G is a homogeneous space.

Lemma. There exists to within equivalence a unique quasi-invariant measure on G/H.

Given the above lemma we can define for this quasi-invariant measure the Radon-Nikodym derivative of the translate measure:

$$\rho(g,x)=\frac{d[\mu\cdot g]}{d\mu}(x) \qquad g\in G \quad x\in X=G/H.$$

Recall that $\mu\cdot g$ is defined by $\mu\cdot g(E)=\mu(E\cdot g)$ where E is a subset of G/H.

Lemma. i) $\rho(g,x)$ is a Borel function of G x G/H

ii) $\rho(g_1g_2,x)=\rho(g_1,x)\ \rho(g_2,xg_1)$ for $g_1,g_2\in G$ and $x\in G/H$.

We can now define the induction procedure.

Definition. Let G be a separable locally compact group, H a closed subgroup of G, and γ a representation of H on a Hilbert space $\tilde{\eta}_\gamma$. Let μ be the quasi-invariant Borel measure on G/H. Let $\tilde{\eta}(\gamma)$ be the set of all functions $f:G\longrightarrow\tilde{\eta}_\gamma$ such that

i) $g\longmapsto\langle f(g),\xi\rangle$ is Borel for all $\xi\in\tilde{\eta}_\gamma$

ii) $f(hg)=\gamma(h)f(g)$ $\quad h\in H$ for (Haar) almost all $g\in G$

iii) $\int_{G/H}\|f(g)\|^2\,d\mu(x)<\infty.$

Note that iii) makes sense because i) and ii) imply $\|f(g)\|^2$ is a Borel function constant on each right coset x=Hg. So we can define $\bar{f}(x)=\bar{f}(Hg)=\|f(g)\|^2$ and interpret iii) as

iii') $\int_{G/H} \bar{f}(Hg)\, d\mu(x) < \infty.$

Clearly $\tilde{\mathcal{H}}(\gamma)$ is a separable Hilbert space with inner product

iv) $\langle f_1, f_2 \rangle_{\tilde{\mathcal{H}}(\gamma)} = \int_{G/H} \langle f_1(g), f_2(g) \rangle_{\mathcal{H}_\gamma} d\mu(x)$

where, of course, the usual equivalence up to sets of measure zero is assumed. Again the integral iv) makes sense because ii) implies $\langle f_1(g), f_2(g) \rangle$ is constant on right cosets.

For each $g, g' \in G$ and $f \in \tilde{\mathcal{H}}(\gamma)$, let

$$\pi(g) f(g') = f(g'g) \sqrt{\rho(g, \bar{g}')} \qquad \text{where } g' \longmapsto \bar{g}'$$

is the projection map, i.e. $\bar{g}'$ is the right coset containing g'. It is easy to see that is a unitary representation of the entire group G. We call π the representation induced from H to G by γ and write $\pi = \mathrm{Ind}_H^G \gamma$.

At first it may seem that the quasi-invariant measure plays an essential role in determining the induced representation; but Mackey (1952) shows that induction only depends on the measure's equivalence class which, by the lemma, is unique.

We will be very interested in the so-called <u>monomial representations</u>.

<u>Definition</u>. If γ is a one-dimensional representation of $H \subseteq G$, then $\mathrm{Ind}_H^G \gamma$ is a <u>monomial representation</u>.

<u>Theorem</u> (Mackey 1952). If $\int_\Lambda^\oplus \gamma_\lambda \, d\mu(\lambda)$ is a direct integral decomposition of γ, a representation of the

closed subgroup H of G, then

$$\mathrm{Ind}_H^G = \int_\Lambda^{\oplus} \mathrm{Ind}_H^G \gamma_\lambda \, d\mu(\lambda) .$$

Thus a representation induced from a reducible representation is reducible. However, irreducible representations do not necessarily induce irreducible representations.

3.13 Imprimitivity Theorem

We have now finished describing the building up procedure. Now let us find a projection-valued measure commuting with an induced representation.

Let B be a Borel subset of X=G/H (H closed in G), let B' be the image under the inverse $B' = \rho^{-1}(B)$ where ρ is the natural epi $\rho: G \dashrightarrow G/H$. We obtain a self-adjoint projection $E(B): \tilde{\eta}(\gamma) \longrightarrow \tilde{\eta}(\gamma)$ defined by

$$f(g) \xmapsto{E(B)} I_B(g) \cdot f(g) .$$

It is clear that E(B) is a projection-valued measure. Evidently with $\pi = \mathrm{Ind}_H^G \gamma$ we have

$$\pi(g) E(B) \pi(g^{-1}) = E(B \cdot g^{-1}) .$$

Definition. Let π be a representation of G. (E, Λ) is a system of imprimitivity of π if

1) Λ is a Borel G-space

2) E is a projection-valued measure on Λ such that

$$\pi(g) E(B) \pi(g^{-1}) = E(B \cdot g^{-1})$$

for $g \in G$, B a Borel subset of Λ.

If Λ is homogeneous (i.e. a transitive G-space), then we

say that (E,Λ) is a transitive system of imprimitivity.

The above example shows that an induced representation gives rise to a transitive system of imprimitivity. Mackey (1949, 1958) gives us a theorem that goes the other direction.

Theorem (Imprimitivity). Let π be a unitary representation of G on a Hilbert space $\mathcal{H}_\pi$ and let E be a system of imprimitivity for π based on G/H, H a closed subgroup. Then there exists a representation γ of H, uniquely determined up to equivalence, and a unitary map

$$U:\mathcal{H}_\pi \longrightarrow \mathcal{H}(\gamma)$$

such that for all $g\in G$ and Borel sets $B\subseteq G/H$ we have

$$U\pi(g)=[\mathrm{Ind}_H^G\gamma(g)]U \qquad UE(B)=E\gamma(B)U,$$

where $E_\gamma(B)f(g)=I_B(\bar{g})f(g)$.

3.14 Semidirect Products

We are now very close to obtaining our desired direct integral decomposition. All the groups in which we are interested have a form in which the induction procedure is particularly easy to apply. This form is given in the following

Definition. Let H and N be locally compact groups. If we can find a homomorphism $\alpha:H\longrightarrow \mathrm{Aut}(N)$ satisfying the condition that the map $H\times N\longrightarrow N$ given by $(h,n)\longrightarrow \alpha(h)n$ is continuous, then we can construct the semidirect product $G=H\times_\alpha N$ of H and N.

Let G be the set of all pairs (h,n) and define a group operation as

$$(h_1,n_1)\ (h_2,n_2)=(n_1\ (h_1)\,n_2,h_1h_2),$$

where h_1h_2 signifies the compound operation of h_2 followed by h_1. It is obvious that $G=H\times_{\alpha}N$ is a group and that 1) N is isomorphic to a normal subgroup consisting of all elements (e,n); 2) H is isomorphic to a subgroup with elements (h,e).

The canonical example is a group G=HN where $H\cap N=\{e\}$, N normal. The α can be taken to be conjugation of N by $h\in H$, i.e. $\alpha(h)\,n=hnh^{-1}$.

In the vision groups, G equals $H\times_{\alpha}N$ where N is not only normal but Abelian. Let $G=H\times_{\alpha}N$ be a semidirect product with N Abelian and π be a representation of G. Then restricting π to H and to N we obtain representations $\pi_1=\pi|_H$, $\pi_2=\pi|_N$. Since G is semidirect we have g=hn and

$$\pi(g)=\pi_1(h)\,\pi_2(n)$$

$$\pi_1(h)\,\pi_2(n)\,\pi_1(h^{-1})=\pi_2(\alpha(h)\,n).$$

Now we know from Stone's theorem that $\pi_2=\pi|_N$ can be written

$$\pi_2=\int_B^{\oplus}\chi\ dE(\chi) \qquad B \text{ a Borel subset of } \hat{N}.$$

It is easy to see that E and π_1 satisfy

$$\pi_1(h)\,E(B)\,\pi_1(h^{-1})=E(h\cdot B)$$

where B is a Borel subset of $\hat{N}$ and for $\chi\in\hat{N}$, $[h\cdot\chi](n)=\chi(\alpha(h^{-1})\ n)$.

Now suppose the measure E is concentrated on a single

orbit of $\hat{N}$, say with element χ, i.e. $E(B)=0$ if $B\cap\mathcal{O}_\chi=\phi$. Let H be the stabilizer subgroup of χ in H. So there is a Borel isomorphism $H/H_\chi\simeq\mathcal{O}_\chi$, and $E(\cdot)$ is a system of imprimitivity for π_2 on H/H_χ. By the imprimitivity theorem, we have

$$\pi_2 = \mathrm{Ind}_{H_\chi}^{H}\nu$$

for some representation ν of H .

Definition. G is a regular semidirect product of H and N if N contains a countable family $\{B_1, B_2, \ldots\}$ of Borel sets such that

1) $B_i = \bigcup \mathcal{O}_{\chi_\alpha} \quad \chi_\alpha \in \hat{N}$

2) for every χ, $\mathcal{O}_\chi = \bigcap_{\chi\in B_j} B_j$,

that is, B_i is a union of orbits and each orbit is a countable intersection of the B_i that contain it.

Theorem. Let G be a regular semidirect product of H and N with N Abelian and normal. Write an orbit as $\mathcal{O}_\chi$, $\chi\in\hat{N}$, and isotropy subgroup of χ in H as H_χ under the action $[h\cdot\chi](n)=\chi(\alpha(h)\,n)$. Let π be an irreducible representation with $\pi_1=\pi|_H$, $\pi_2=\pi|_N$. Then the projection-valued measure defined by π_2 in $\hat{N}$ is concentrated in a single orbit $\mathcal{O}_\chi$ and $\pi_1=\mathrm{Ind}_{H_\chi}^{H}\nu$ for some ν, an irreducible representation of H. Every triple $\mathcal{O}_\chi, H_\chi, \nu$ arises in this way and $\mathcal{O}_{\chi'}, H_{\chi'}, \nu'$ is equivalent if $\mathcal{O}_\chi=\mathcal{O}_{\chi'}$, ν is equivalent to ν'. Now if G is a regular semidirect product, then each of the Borel sets B_i is invariant under G-action. Thus the σ-algebra Λ

generated by $\{B_i\}$ is also invariant under G-action.

The system of imprimitivity $(E,\hat{N})$ restricted to Λ, the subsigma algebra, is not only a system of imprimitivity but also a commuting projection-valued measure. Mackey's theorem, in conjunction with Mautner's, says that the Boolean algebra generated by this projection-valued measure is maximal.

Fundamental Structure Theorem. Let π be a unitary representation on a Hilbert space $\tilde{\mathcal{H}}$. Let $A\in\mathcal{R}(\pi,\pi)$ be a normal operator. Then there is a spectral resolution of A

$$A=\int_{\Lambda} a(\lambda)\,dE(\lambda)$$

corresponding to a direct integral decomposition

$$\tilde{\mathcal{H}} = \int_{\Lambda}^{\oplus} \tilde{\mathcal{H}}_\lambda \, d\mu(\lambda)$$

where

$$\pi = \int_{\Lambda}^{\oplus} \pi_\lambda \, d\mu(\lambda)$$

and the π_λ are μ-almost everywhere irreducible.

Proof (Sketch). By Mautner's theorem the maximal Boolean algebra of projections Q give rise to the latter direct integral decomposition with the stated properties. Now let

$$A=\int_{\sigma(A)} \zeta \quad dE'(\zeta)$$

be the spectral resolution of A arising from the classical spectral theorem. Since A commutes with π, so does every operator in the commutative C^*-algebra generated by $\{A,A^*\}$. In particular, the projections E'(B') for every

$B' \subseteq \sigma(A)$ commute with π. Now fix a specific $E'(B')$. Since $E'(B)$ is a projection that commutes with π and Q, the Mautner Boolean algebra of projections, is maximal, $E'(B') \in Q$. This shows that

$$E'(B') = \int_{\Lambda} I_C(\lambda)\, d\mu(\lambda)$$

for some $C \subseteq \Lambda$. (That C is Borel is a true but highly involved result. We shall refer the interested reader to Mackey (1957)).

Thus by constructing the usual sequence of simple functions, one can obtain an isomorphism

$$A = \int_{\Lambda} a(\lambda)\, d\mu(\lambda).$$

Let E be the projection-valued measure corresponding to the direct integral decomposition. The theorem follows.

This then gives the result we have been looking for. To use it, we will assume the operator A is the visual system. We apply the theorem using a direct integral decomposition corresponding to representations of the vision groups VG_i. Note that although $a(\lambda)$ is not given to us, the structure of the operator -- given by the projection-valued measure E -- follows only from knowledge of the symmetry group and the theorem.

3.15 Examples

Our first example is VG_1, the ax+b group, with the representation defined on $L^2(R, \mu)$ by

$$\pi(a,b) : f(x) \longmapsto a^{\frac{1}{2}} f(ax+b).$$

G=HN is the semidirect product of translations (N) and dilatations (H). $\pi_1=\pi|_N$ is just the additive group of the real line and $\hat{N}$ is the space of familiar Fourier characters. Notice that scaling the space R by a corresponds to scaling the character by 1/a. Since a is restricted to the positive reals, the positive characters always are scaled into themselves. Also any positive character can be scaled onto any other positive character. The likewise situation holds for the negative characters. Also the DC character is unchanged by any scaling operation. Thus there are just three orbits $\mathcal{O}_+, \mathcal{O}_-, \mathcal{O}_0$ where $\mathcal{O}_+$ contains all characters of the form $\chi(b)=e^{i\omega b}$, $\omega>0$. Similarly, $\mathcal{O}_-$ and $\mathcal{O}_0$ contain the negative frequencies and DC term.

Thus the fundamental structure theorem gives us that the structure for any operator in $\mathcal{R}(\pi)$ is

$$A=a_-E_-+a_0E_0+a_+E_+$$

where E_-, E_0, E_+ are the projections onto the negative, DC and positive frequency subspaces of the direct integral decomposition. Any commuting operator A can be specified with three numbers (a_-,a_0,a_+).

Example 2 gives our first result about vision. Let $G=VG_2$ be the group of translations (b_1,b_2) and dilatation $a>0$ in the plane. Let π be defined on $L^2(R^2,\mu)$ by

$$\pi(a,b_1,b_2):f(x_1,x_2)\longmapsto af(ax_1+b_1,ax_2+b_2).$$

G=HN is the semidirect product of translations (N) and

dilatations (H). $\pi_1 = \pi|_N$ is just the additive group of the plane and $\hat{N}$ is the space of Fourier characters. Any character can be taken by a dilatation onto any other along the ray connecting the origin with that character. So there are many orbits, each corresponding to an angle θ, $0 \leq \theta < 2$, where $\mathcal{O}_\theta$ contains all characters of the form

$$\chi(b_1, b_2) = e^{ir(\cos\theta \cdot b_1 + \sin\theta \cdot b_2)} \quad , \; r > 0.$$

Thus the structure for any operator in $\mathcal{R}(\pi)$ is given by the structure theorem to be

$$A = \int_{(H)} a(\theta)\, dE(\theta)$$

where E(B) is the projection-valued measure defined on the countable number of "wedge" sets in $\hat{N}$ (see Figure 3.1), where θ, θ' are rational.

These wedge set projections are familiar to vision theorists (Pantle and Sekuler 1968, Blakemore and Campbell 1969). It is interesting that symmmetry considerations alone can predict their existence.

Next, let $G = VG_3$, the group of rigid motions of the plane. Let us write a rotation of θ by

$$r(\theta) : (x_1, x_2) \longmapsto (r_1(\theta), r_2(\theta)).$$

Then π is defined by

$$\pi(\theta, b_1, b_2) : f(x_1, x_2) \longmapsto f(r_1(\theta) + b_1, r_2(\theta) + b_2).$$

Again, G is a semidirect product, with $\hat{N}$ the additive group of the plane. For each $r \geq 0$, there is an orbit

$$\mathcal{O}_r = \{\chi \mid \chi(b_1, b_2) = e^{ir(\cos\theta \cdot b_1 + \sin\theta \cdot b_2)} \qquad 0 \leq \theta < 2\pi\}.$$

Thus the structure for any operator in $\mathcal{R}(\pi)$ is

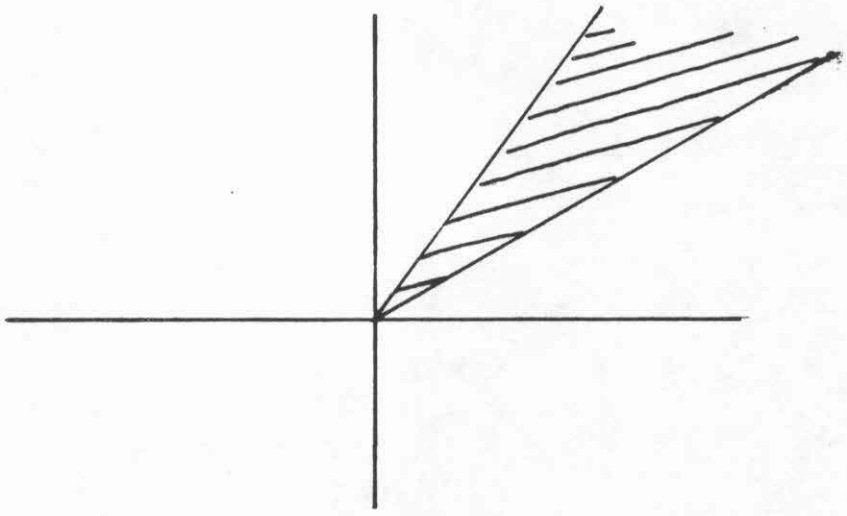


Figure 3.1 - Wedge sets

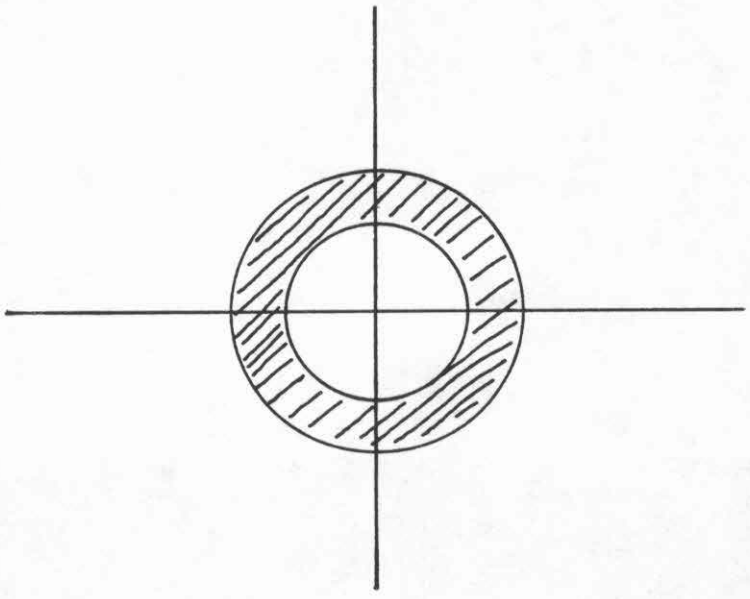


Figure 3.2 - Ring sets.

$$A=\int_R a(r)\,dE(r)$$

where E(B) is the projection-valued measure defined on the countable number of "ring" sets (see Figure 3.2), where r, r' are rational. Note that to find the determining function a(r), one need only test the operator at a single orientation. This, of course, is exactly what psychophysicists do.

The fourth example $G=VG_4$ points out clearly a difficulty with our structure theory; one that is implicit in the previous three examples. Using the notation defined previously, we define π by

$$\pi(a,\theta,b_1,b_2):f(x_1,x_2)\longmapsto af(ar_1(\theta)+b_1,ar_2(\theta)+b_2).$$

The semidirect product structure with N being the translations is apparent. It is identical to our canonical example of a semidirect product. There are only two orbits

$$\mathcal{O}_0=\{\chi(b_1,b_2)\equiv 1\}\qquad \mathcal{O}_1=\{\chi|\chi(b_1,b_2)\not\equiv 1\}.$$

Therefore the structure of any operator in $\mathcal{R}(\pi)$ is

$$A=a_0E_0+a_1E_1$$

where E_0 is the "DC" projection operator, i.e. the McCulloch-Pitts (1947) group average, and E_1 is the "everything else" projection.

Clearly something is wrong here. Form perception is symmetric under VG_4, yet the structure theorem gives only trivial structure to operations in $\mathcal{R}(\pi)$. Of course, one of our assumptions must be false. The first assumption is

that the form operator is linear. The second is that the form operator is contained in $\mathcal{R}(\pi)$. We remove these assumptions in the sequel. The first steps toward a generalization to nonlinear intertwining operators is taken in Chapter 4. As is to be expected, much less is known about nonlinear operators, but it is necessary to study them since the eye is clearly nonlinear. Limitations to the second assumption will be relaxed in the sequel by considering linear operators not in $\mathcal{R}(\pi)$ but in $\mathcal{R}(\pi,\pi\otimes\pi)$, where $\pi\otimes\pi$ is the tensor product representation to be defined below. To do this we will use a particularly simple version of the nonlinear theory (which works of course for linear operators too).

CHAPTER 4

NONLINEAR SYMMETRY THEORY

4.1 Introduction

This chapter will develop a new symmetry theory which does not assume that the unknown operator is linear. In discarding the linear assumption, we incur two disadvantages. First, the theory does not go as far: that is, a structure theorem that gives us a way to map functions $a(\lambda)$ into symmetric operators is beyond our efforts here. Second, we must assume an a priori operator structure, albeit a very general structure such as

$$f(x)=\int_{-\infty}^{\infty}k(x,\xi,f(\xi))\,d\xi.$$

This requirement is not as severe as in the classical case because the operator structure is so general that ad hoc modelling is avoided.

We shall in the next section demonstrate the basic idea of this chapter in a straightforward example, which is interesting in itself and forms a basis for the discussion in Chapter 5. Section 4.3 will give basic definitions which will be used in obtaining the actual structure theory presented in Section 4.4. Arguments then will be given to show how a system of partial differential

equations may be obtained whose general solution gives the kernel for an operator that obeys the symmetry group. Finally, examples will be given illustrating these new methods.

4.2 An Intertwining Operator

Let us illustrate the basic aim of this chapter by calculating an intertwining operator of a certain form between a representation π and its tensor square $\pi\otimes\pi$, where π is a (non-unitary) representation of the ax+b group given on $L^2(R,\mu)$ by

$$\pi(a,b):f(x)\longmapsto f(ax+b)$$

and $\pi\otimes\pi$ is the representation on $L^2(R,\mu)\otimes L^2(R,\mu)\simeq L^2(R^2,\mu)$ given by

$$\pi\otimes\pi(a,b):f(x,y)\longmapsto f(ax+b,ay+b).$$

The form of the intertwining operator A will be chosen to be the very common Fredholm operator:

$$\hat{f}(x,y)=\Phi f(x,y)=\int_{-\infty}^{\infty}K(x,y,\xi)\,f(\xi)\,d\xi.$$

Now we can proceed. From the definition of intertwining we have

$$(\Phi\pi)\,f=\pi\otimes\pi\,(\Phi f)$$

or

$$\int_{-\infty}^{\infty}K(x,y,\xi)\,f(a\xi+b)\,d\xi=\int_{-\infty}^{\infty}K(ax+b,ay+b,\xi)\,f(\xi)\,d\xi.$$

Substituting ξ for $a\xi+b$ in the LHS, we obtain

$$\int_{-\infty}^{\infty}K(x,y,\tfrac{\xi-b}{a})\,f(\xi)\,\frac{d\xi}{|a|}=\int_{-\infty}^{\infty}K(ax+b,ay+b,\xi)\,f(\xi)\,d\xi$$

or

$$\int_{-\infty}^{\infty}\{\tfrac{1}{|a|}\,K(x,y,\tfrac{\xi-b}{a})-K(ax+b,ay+b,\xi)\}\,f(\xi)\,d\xi=0$$

for all $f(\xi)\in L^2(R,\mu)$; thus

$$\frac{1}{|a|}K(x,y,\frac{\xi-b}{a})=K(ax+b,ay+b,\xi) \qquad \mu.a.e.$$

Making the substitution $a\xi+b$ for ξ we obtain

$$K(x,y,\xi)=|a|\,K(ax+b,ay+b,a\xi+b).$$

Since this equation holds for $a\neq 0$, b arbitrary real numbers, we may treat the above equation as a functional equation. Letting $a=\frac{1}{x-\xi}$, $b=-\frac{\xi}{x-\xi}$, we obtain

$$K(x,y,\xi)=\frac{1}{|x-\xi|}K(1,\frac{y-\xi}{x-\xi},0).$$

If this step makes the reader uneasy, then by affixing subscripts to the x, y, ξ, we may treat this as an ordinary equation involving the quantity $K(x_0,y_0,\xi_0)$. (Note however that this is a purely formal manipulation; we have not stipulated what happens in the case $\xi=x$ or $\xi=y$.)

Renaming the function $K(1,\zeta,0)=k(\zeta)$, we have as the intertwining operator

$$\hat{f}(x,y)=\Phi f(x,y)=\int_{-\infty}^{\infty}k(\frac{y-\xi}{x-\xi})\,f(\xi)\,\frac{d\xi}{|x-\xi|}\ .$$

These computations have been purely formal, as in freshman calculus; in particular, the meaning of the above integral has not been established. These points will be treated in Chapter 5. Readers interested only in the Mandala transform may safely skip directly to Chapter 5.

This example shows the salient features of the nonlinear structure theory (which works of course for linear operators also). The inputs to this new structure theory are the action of some transformation group and an

a priori operator structure. We have said action instead of representation not only because a group element may be taken into a nonunitary transformation (as in the above example), but also because a group element may be taken into a nonlinear transformation (as in VG_5 of Section 2.3).

The new technique we present below is considerably more abstract than the above example. But it more than repays the loss in concreteness by a gain in applicability. With this new method it is possible to determine in a uniform, straightforward way the structure of operators arising from a priori structures of much greater complexity and under more complicated groups and group actions. This provides a key procedure in the formal modelling approach to visual system behavior.

4.3 Definitions

First we will need some definitions from infinite-dimensional manifold theory. This theory provides the most natural expression for the key equation we derive in Section 4.4 which we have called the fundamental symmetry formula. The primary reference for infinite-dimensional manifold theory is Lang (1962), which our definitions follow closely.

Let Φ be a map between two Banach spaces E,F, where $\mathrm{dom}\Phi \subseteq U$, U open in E:

$$\Phi : U \longrightarrow F.$$

Then Φ is (Frechet) differentiable at $x_0 \in U$ if there exists a continuous linear map λ of E into F such that

$$\Phi(x_0+y) = \Phi(x_0) + \lambda y + o(y)$$

where $o(y)$ is a map from $E \longrightarrow F$ such that

$$\lim_{\|y\| \to 0} \frac{\|o(y)\|_F}{\|y\|_E} = 0.$$

If Φ is differentiable, then we say that λ is the derivative of Φ at x_0 and write λ as $\Phi'(x_0)$. If Φ is differentiable at every $x_0 \in U$, then Φ' is a map

$$\Phi' : U \longrightarrow L(E,F)$$

where L(E,F) is the space of bounded linear operators from E to F (with the uniform norm).

Let X be Hausdorff. An atlas on X is a collection of pairs (U_i, φ_i) such that

1) each U_i is an open subset of X and the U_i cover X;
2) Each φ_i is a homeomorphism of U_i onto an open set $\varphi_i U_i$ of some Banach space E and for any i,j, $\varphi_i(U_i \cap U_j)$ is open in E;
3) the map $\varphi_j \varphi_i^{-1} : \varphi_i(U_i \cap U_j) \longrightarrow \varphi_j(U_i \cap U_j)$ is a C^∞ diffeomorphism.

Each (U_i, φ_i) is called a chart of the atlas and each φ_i a coordinate map. If $x \in X$ is such that $x \in U_i$, then (U_i, φ_i) is said to be a chart at x. Let (U, φ) be some pair with U open in X and $\varphi : U \longrightarrow \varphi U \subseteq E$ a homeomorphism. Then (U, φ) is said to be compatible with the atlas $\{(U_i, \varphi_i)\}$ if each $\varphi_i \varphi^{-1}$ satisfies condition 3). When X is a Hausdorff space with an atlas $\{(U_i, \varphi_i)\}$ on E, then X is a (Banach)

manifold (or an E-manifold) if the atlas is maximal with respect to compatibility, i.e. if (U,φ) is compatible with the atlas $\{(U_i,\varphi_i)\}$, then $(U,\varphi)\in\{(U_i,\varphi_i)\}$.

Let X, Y be two manifolds. Let $\Phi: X \dashrightarrow Y$ be a map. We say that Φ is C^p if for atlases $\{(U_i,\varphi_i)\}$ of X and $\{(V_j,\psi_j)\}$ of Y, we have

$$\psi_j \Phi \varphi_i^{-1}: \varphi_i U_i \longrightarrow \psi_j V_j$$

is a C^p map.

For $x\in X$, X an E-manifold, consider the triples (U,φ,v) where (U,φ) is a chart at x and $v\in E$. We say that two triples (U,φ,v), (V,ψ,w) are equivalent if the derivative of $\psi\varphi^{-1}$ at φx maps v onto w, i.e.

$$(\psi\varphi^{-1})'(\varphi x)\, v = w.$$

This is an equivalence relation by the chain rule. The set of equivaluence classes of such triples forms the tangent space of X at x, written X_x; each equivalence class represented by a triple is called a tangent vector of X at x. Note that for each chart (U,φ) we can consider a mapping of the tangent vector represented by (U,φ,v) onto E given by $(U,\varphi,v)\longmapsto v$. It is clear that this mapping is well-defined, and that we can pull back the topology on E to X (namely the weakest topology making the mapping continuous). Now if $\Phi: X \longrightarrow Y$ is a smooth mapping from an E-manifold X to an F-manifold Y, then we can define a mapping of tangent spaces

$$\Phi_* : X_x \longrightarrow Y_{\Phi(x)}$$

called the differential of $\underline{\Phi}$ at x, the prolongation of $\underline{\Phi}$ at x, or the tangent map of $\underline{\Phi}$ at x by

$$\underline{\Phi}_*: (U,\varphi,v) \longmapsto (V,\psi,w)$$

where (V,ψ) is any chart at $\underline{\Phi}(x)$ and

$$(\psi\underline{\Phi}\varphi^{-1})'(\varphi x)\, v = w.$$

A vector bundle B on X is a tuple $(B,X,\pi,F,\{U_i,\tau_i\})$ such that

1) B is a topological space (the bundle space)

2) X is an E-manifold (the base space)

3) π is a surjective mapping $\pi: B \longrightarrow X$ (the projection)

4) for each $x \in X$, $\pi^{-1}(x)$ is a Banach space written B_x (the fiber at x)

5) F is a Banach space (the fiber)

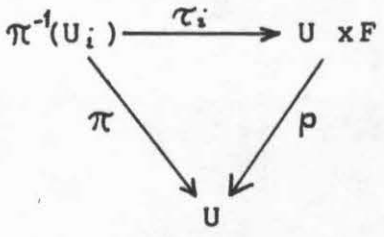
6) $\{U_i\}$ is an open covering of the base space X (the trivializing covering)

7) for each U_i there exists a τ_i such that

$$\tau_i: \pi^{-1}(U_i) \longrightarrow U_i \times F$$

is a homeomorphism and isomorphism

8) τ_i commutes with the projections



where $p: (x,v) \longmapsto x$

9) the restriction $(\tau_i)_x: B_x \longrightarrow \{x\} \times F$ is an

isomorphism

10) if $U_i, U_j \in \{U_i\}$, then the map τ_{ij} of $U_i \cap U_j$ into $L(E,E)$ given by

$$\tau_{ij}(x) = (\tau_i)_x \, [(\tau_j)_x]^{-1}$$

for $x \in U \cap V$ is smooth. (τ_{ij} is the transition mapping.)

The tangent bundle TX of an E-manifold X is the disjoint union of the tangent spaces, i.e.

$$TX = \coprod_{x \in X} X_x$$

with the following vector bundle structure. Let $\pi : TX \longrightarrow X$ be the natural projection $\pi : (U, \varphi, v) \longmapsto x$ where (U, φ, v) represents a tangent vector at x. From the definition of tangent spaces as equivalence classes of triples (U, φ, v), we get immediately a bijection

$$\tau_U : \pi^{-1}(U) \longrightarrow U \times E.$$

If (U_i, φ_i), (U_j, φ_j) are two charts and φ_{ji} is the map $\varphi_j \varphi_i^{-1}$ defined on $\varphi_i(U_i \cap U_j)$, then we can obtain a transition mapping

$$\tau_{ji} = (\tau_j \, \tau_i^{-1}) : \varphi_i(U_i \cap U_j) \times E \longrightarrow \varphi_j(U_i \cap U_j) \times E$$

by

$$\tau_{ji}(x,v) = (\varphi_{ji}(x), (\varphi_{ji})'(x)\, v) \quad \text{for } x \in U_i \cap U_j \text{ and } v \in E.$$

Since the derivative of φ_{ji} is linear and an isomorphism (by definition of coordinate maps), we have that TX is a vector bundle.

A vector field is a cross-section of the tangent bundle, i.e. a smooth mapping

$$\xi : X \dashrightarrow TX.$$

That is, $\pi \circ \xi = id$.

Given a chart (U,φ) and a trivializing map $\tau_u : \pi^{-1}(U) \longrightarrow U \times E$, we can express the vector field restricted to U as a pair of functions

$$(i, \xi_u) = \tau_u \circ \xi,$$

i.e.

$$\tau_u : x \longmapsto (x, \xi_u(x)) \qquad \text{where } x \in U,\ \xi_u(x) \in E.$$

The map $\xi_u = p_2 \circ \tau_u \circ \xi : U \longrightarrow E$ is called the <u>principal part</u> of ξ with respect to U.

Given a mapping $\Phi : X \longrightarrow Y$ and vector fields ξ, η on X and Y respectively, we say that ξ and η are <u>Φ-related</u> if the following holds:

$$\Phi_x(x)\,\xi(x) = \eta(\Phi(x)) \qquad \text{for all } x \in X,$$

where Φ_* is the differential of Φ at x. If ξ and η are Φ-related vector fields, then we sometimes write

$$\Phi \cdot \xi = \eta \text{ or } \xi^{\Phi} = \eta.$$

Note that given an arbitrary smooth mapping, we may not find a Φ-related vector field η for a given vector field ξ, e.g. Φ may not be one-to-one and $\xi(x) \neq \xi(x')$ for x, x' such that $\Phi x = \Phi x'$. If, however, Φ is a diffeomorphism, then a Φ-related vector field may always be found.

A <u>Lie group</u> is a finite-dimensional manifold G with a group operation such that the group operation given by $(a,b) \longmapsto ab^{-1}$ is a smooth mapping from $G \times G \longrightarrow G$, where $G \times G$ is the product manifold.

The maps L_a, R_a of G onto G, the left and right translations, given by

$$L_a(b)=ab \qquad R_a(b)=ba,$$

are by definition diffeomorphisms. Thus for every vector field ξ on G we can find an L_a-related vector field ξ', i.e.

$$\xi'=L_a\cdot\xi .$$

If $\xi'=\xi$, then we say that ξ is <u>left invariant</u>. The set of all left invariant vector fields of G is called the <u>Lie algebra</u> $\mathcal{G}$ of G.

A well-known correspondence between the Lie algebra $\mathcal{G}$ and the tangent space at e, G_e, is given by

$$\xi \longmapsto \xi(e) \qquad \xi\in\mathcal{G},\ \xi(e)\in G_e$$

$$t\longmapsto \xi \quad \text{where } \xi(g)=(L_g)_* t .$$

For a G-manifold X we now construct a mapping from $\mathcal{G}$ into the set of vector fields on X.

For a point $x\in X$ choose a chart (U,φ) at x. We may choose an open neighborhood $W\subseteq G$ of the identity $e\in G$ such that

$$\{g\cdot x \mid g\in W\}\subseteq U$$

since the action of G on X is continuous. Thus it is possible to construct a <u>smooth</u> mapping $\tilde{\varphi}_x : W\longrightarrow E$ given by

$$\tilde{\varphi}_x : g\longrightarrow \varphi(gx) .$$

Now we may define a mapping from the tangent space of G at e given by the tangent map at e

$$(\tilde{\varphi}_x)_* : G_e\longrightarrow E.$$

Now let $v=(\tilde{\varphi}_x)_* t$ where $v\in E$, $t\in G_e$. By the chain rule, this

construction induces equivalent triples $(U,\varphi,(\tilde{\varphi}_x)_* t)$, $(V,\psi,(\tilde{\psi}_x)_* t)$. Thus, for every $x\in X$, we have defined a mapping from G_e, the tangent space of G at e, to X_x, the tangent space of X at x.

It follows from the definitions that this mapping takes G_e into the set of smooth vector fields of X. Given the correspondence between G_e and $\mathcal{G}$, we may define a mapping from $\mathcal{G}$ to the left-invariant vector fields in X, written

$$\xi \longmapsto \bar{\xi}_X$$

where $\xi\in\mathcal{G}$ and $\bar{\xi}_X$ is a vector field of X.

4.4 Structure Theory

An analog of an intertwining operator for the nonlinear theory is given by the next definition.
An <u>equivariant</u> map Φ of two G-spaces X,Y is a map such that

$$g\Phi(x)=\Phi(gx) \qquad \text{for } x\in X.$$

Note that if the G-action is given by linear representations on Hilbert spaces X,Y, then this definition collapses to that of an intertwining operator.

For much of our purposes we will not need to use the full restriction of equivariance, but an alternative.

<u>Definition</u>. If X,Y are manifolds possessing smooth G-actions, then $\Phi:X\longrightarrow Y$ is said to be <u>infinitesmally equivariant</u> if $\Phi\cdot\bar{\xi}_x=\bar{\xi}_Y$ for $\xi\in\mathcal{G}$.

It is obvious that equivariance implies infinitesmal equivariance. The converse is a much more delicate question, beyond the scope of the present discussion; see Palais (1962). For this reason, we shall now work exclusively with the necessary condition of infinitesmal equivariance. It is equivariance, though, that is the primary concept. Thus the reader should keep in mind that what we are searching for in the structure theory is the structure of equivariant operators as revealed by its infinitesmal necessary condition.

We would like to recast the above definition of infinitesmal equivariance so that computations may be more readily performed. Let $\Phi: X \longrightarrow Y$ be a smooth mapping. Fix a coordinate chart (U,φ) at x and suppose $\Phi(x) \in V$ for (V,ψ), a chart at $\Phi(x)$. Then by definition Φ is infinitesmally equivariant if

$$\Phi \cdot \bar{\xi}_X = \bar{\xi}_Y$$

or at $x \in X$

$$\Phi_* \bar{\xi}_X(x) = \bar{\xi}_Y(\Phi(x)) \qquad \text{where } \bar{\xi}_X(x) \in X_x.$$

Taking principal parts we have

$$\boxed{\Phi'(x)(\bar{\xi}_X(x))_U = (\bar{\xi}_Y(\Phi(x)))_V .}$$

This result we shall call the <u>fundamental symmetry formula</u>.

4.5 Examples

We shall now show how the fundamental symmetry formula may be used in conjunction with general a priori models of vision to obtain a system of first-order partial differential equations. These PDE will then be solved to give us the structure of the operator kernels in our a priori kernel.

We now present the procedure by spelling out a recipe:

1) Determine the symmetry Lie group G and its actions on the domain and range manifolds.
2) Calculate the Lie algebra $\mathcal{G}$ of G and determine a basis $\xi^1,\ldots,\xi^n$.
3) Calculate the infinitesmal actions of the ξ^i; i.e. calculate $\bar{\xi}^i_X, \bar{\xi}^i_Y$ and their principal parts $(\bar{\xi}^i_X)_u$, $(\bar{\xi}^i_Y)_v$.

These first three steps are, of course, independent of any a priori operator structure and can be performed once and for all given just the symmetry group and its action.

4) Choose an a priori operator structure and apply the fundamental symmetry formula.
5) Deduce the system of first order PDE for the kernel.
6) Solve the PDE for the symmetric system kernel.

Let the manifolds X,Y be spaces of one-dimensional images $\{f(x) \mid f \in W^{\infty,2}(R), f>0\}$, where $W^{\infty,2}$ is the Sobolev

space with norm $\|f\| = (\int \sum |D^\alpha f|^2 \, dx)^{\frac{1}{2}}$. We have changed the notation slightly by letting the elements of X,Y be denoted not by x but by f(x), thus saving x for the underlying space R. In Step 1 we let the symmetry group action on X be given by VG and VG , i.e.

$$g_{y,p} \cdot f(x) = [f(x+y)]^p \qquad \text{for } y \in R,\ p \geq 0,\ f \in X.$$

Clearly we can take U=X and $\varphi = id$, where $\varphi(U) = X$ is an open subset of the Sobolev space (which is Banach) $W^{\infty,2}$. We have

$$\tilde{\varphi}_f : g_{y,p} \longmapsto [f(x+y)]^p .$$

Let the symmetry group action on Y=X be given by

$$g_{y,p} \cdot f(x) = pf(x+y) \qquad y \in R, p \geq 0, f \in Y.$$

Again we may use the chart V=Y and $\psi = id$. We have

$$\tilde{\psi}_f : g_{y,p} \longmapsto p \cdot f(x+y) .$$

For Step 2 we let $\bar{\xi}^1 \in \mathcal{O}_f$ be the vector field corresponding to the one-parameter subgroup $g_{y,1}$ and $\xi^2 \in \mathcal{O}_f$ be the vector field corresponding to the one-parameter subgroup $g_{o,p}$. These obviously form a basis for $\mathcal{O}_f$.

We implement Step 3 by using the definition of infinitesmal action given at the end of Section 4.3. By definition we have

$$(\bar{\xi}^1_x)_u (f) = (\tilde{\varphi}_f)_* \, \xi^1(e) = (\tilde{\varphi}_f)' \, \xi_1(e) .$$

Calculating the last quantity

$$\tilde{\varphi}_f (g_{o,1} + g_{h,1}) = f(x+h) = f(x) + hf'(x) + o(h);$$

thus $(\bar{\xi}^1_x)_u = (\bar{\xi}^1_Y)_V = f' = \frac{df}{dx}$.

Similarly $(\bar{\xi}^2_x)_u = f \cdot \ln f$ and $(\bar{\xi}^2_Y)_V = f$. Table 1 gives the

TABLE 1
GROUP ACTIONS

	Action on $\mathbb{R}^1$	$(\bar{\xi}_{\mathbb{R}^1})_u$	Action on $\mathbb{R}^2$	$(\bar{\xi}_{\mathbb{R}^2})_u$
Translation	$f(x) \mapsto f(x+\alpha)$	$\frac{df}{dx}$	$f(x,y) \mapsto f(x+\alpha, y)$ $f(x,y) \mapsto f(x, y+\alpha)$	$\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$
Dilatation	$f(x) \mapsto f(\alpha x)$	$x \frac{df}{dx}$	$f(x,y) \mapsto f(\alpha x, \alpha y)$	$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y}$
Rotation	—	—	$f(x,y) \mapsto f(x\cos\theta - y\sin\theta,$ $y\sin\theta + x\cos\theta)$	$y\frac{\partial f}{\partial x} - x\frac{\partial f}{\partial y}$
Shear	—	—	$f(x,y) \mapsto f(x+\alpha y, y)$ $f(x,y) \mapsto f(x, y+\alpha x)$	$y\frac{\partial f}{\partial x}$ $x\frac{\partial f}{\partial y}$
Scale	$f(x) \mapsto \alpha f(x)$	f	$f(x,y) \mapsto \alpha f(x,y)$	f
Gamma	$f(x) \mapsto f(x)^\alpha$	$f \ln f$	$f(x,y) \mapsto f(x,y)^\alpha$	$f \ln f$
Bias	$f(x) \mapsto f(x)+\alpha$	1	$f(x,y) \mapsto f(x,y)+\alpha$	1

vector fields corresponding to various symmetries.

For Step 4, let the a priori operator structure be chosen as

$$\Phi f(x)=\int_{-\infty}^{\infty} K(x,\zeta,f(\zeta))\,d\zeta.$$

This is a typical general nonlinear operator. Now, using the fundamental symmetry formula, we have that Φ must satisfy

(4.5.1) $$\Phi'(f)f'=[\Phi(f)]'$$

(4.5.2) $$\Phi'(f)[f\cdot \ln f]=\Phi(f).$$

Up until now we have not used the a priori operator form, so that the above equations hold for all equivariant operators.

Proceeding with the computation of Step 5, we focus on the particular a priori operator structure chosen. Calculating the Frechet derivative we obtain

$$\Phi'(f)h=\int_{-\infty}^{\infty}\frac{\partial K}{\partial f}(x,\zeta,f(\zeta))h(\zeta)\,d\zeta.$$

Thus equations (4.5.1) and (4.5.2) now read

$$\int_{-\infty}^{\infty}\frac{\partial K}{\partial f}(x,\zeta,f(\zeta))f'\,d\zeta=\{\int_{-\infty}^{\infty}K(x,\zeta,f(\zeta))\,d\zeta\}'$$

$$\int_{-\infty}^{\infty}\frac{\partial K}{\partial f}(x,\zeta,f(\zeta))f\ \ln f\ d\zeta=\int_{-\infty}^{\infty}K(x,\zeta,f(\zeta))\,d\zeta.$$

Taking the derivative in the RHS of (4.5.1), we obtain

$$\int_{-\infty}^{\infty}\frac{\partial K}{\partial f}(x,\zeta,f(\zeta))f'(\zeta)\,d\zeta=\int_{-\infty}^{\infty}\frac{\partial K}{\partial x}(x,\zeta,f(\zeta))\,d\zeta.$$

Now the integrand on the LHS is expressible by the chain rule as

$$\frac{\partial K}{\partial f}f'=\frac{d}{d\zeta}\{K(x,\zeta,f(\zeta))\}-\frac{\partial K}{\partial \zeta}.$$

So our formula becomes

$$[K]_{-\infty}^{\infty}-\int_{-\infty}^{\infty}\frac{\partial K}{\partial \zeta}(x,\zeta,f(\zeta))\,d\zeta=\int_{-\infty}^{\infty}\frac{\partial K}{\partial x}(x,\zeta,f(\zeta))\,d\zeta.$$

The first term vanishes so that we have

$$\int_{-\infty}^{\infty} \frac{\partial K}{\partial \zeta} + \frac{\partial K}{\partial x} d\zeta = 0$$

or

$$\frac{\partial K}{\partial \zeta} + \frac{\partial K}{\partial x} = \frac{dG}{d\zeta} \qquad \text{where G vanishes at infinity.}$$

Separating the homogeneous and particular solutions, we can see that the particular solution does not contribute to the final expression; thus we are left with the equation

$$\frac{\partial K}{\partial \zeta} + \frac{\partial K}{\partial x} = 0.$$

Similarly (4.5.2) will give us the equation

$$f \ln f \frac{\partial K}{\partial f} = K.$$

This gives us a system of first-order partial differential equations for the kernel K. We shall now solve this system (Step 6). The characteristic equations for the second equation are

$$\frac{dx}{0} = \frac{d\zeta}{0} = \frac{df}{f \ln f} = \frac{dK}{K}.$$

The independent integrals are

$$x = \alpha$$

$$\zeta = \beta$$

$$\frac{K}{\ln f} = \gamma.$$

Thus the general solution is

$$F(\alpha, \beta, \gamma) \quad F(x, \zeta, \frac{K}{\ln f}) = 0,$$

where F is an arbitrary function. Letting $\gamma = \bar{G}(\alpha, \beta)$ for some arbitrary $\bar{G}$ we obtain

$$\frac{K}{\ln f} = \gamma = \bar{G}(\alpha, \beta) = \bar{G}(x, \zeta)$$

or

$$K(x,\zeta,f)=\bar{G}(x,\zeta)\ln f.$$

The first equation may now be used to calculate $\bar{G}$. It is easy to see that the equation gives

$$\bar{G}(x,\zeta)=k(x-\zeta)$$

for some arbitrary k. Thus we have as a structure for our symmetric transform:

$$\bar{\Phi} f=\int_{-\infty}^{\infty} K(x,\zeta,f(\zeta))\,d\zeta\equiv\int_{-\infty}^{\infty} k(x-\zeta)\ln[f(\zeta)]\,d\zeta.$$

The reader will notice that this form is the well-known homomorphic model for human vision (Stockham 1972, Baudelaire 1973). Note that contrary to the usual way of positing such a model, we have deduced it using only the very general a priori model structure and symmetry considerations.

CHAPTER 5

THE MANDALA TRANSFORM

5.1 Introduction

This chapter will present a transform that arises from the symmetry theories discussed above. Section 2 will define the Mandala transform using the results obtained in Section 4.2. The meaning of this transform will be discussed in the following section by giving analogies to well-known operators, some of which are familiar to engineers acquainted with the signal processing discipline, others of which are familiar to vision theorists as models for brightness perception. Section 4 will develop an inverse, or more accurately, a retract, of the Mandala transform. Finally, the last section will discuss an important approximation practice in signal processing, why it leads to many difficulties, and how the Mandala transform circumvents these difficulties.

Recall that in Section 3.14 we deduced the structure for an intertwining operator for VG_1, the ax+b group. In this case, the structure of the intertwining operators is extremely simple: any operator can be described by a point

in $\mathbb{C}^3$. Similarly for the group of translations, rotations, and dilatations (VG_4), the intertwining operators form a two-dimensional complex vector space. Both of these groups consist of the homogeneous affine transformations of their respective domains. The corresponding structures for intertwining operators are extremely simple, indeed almost trivial. The two situations are so similar that we will deal only with VG_1 from now on. This triviality presents a problem in that the visual system is highly non-trivial but is, at least for form perception, symmetric under the above groups. Thus either the visual system is non-linear and/or it does not intertwine the given representations with themselves. In either case, the methods of Chapter 4 which require neither of these assumptions come to our aid. Examination of experimental evidence in the next chapter seems to show that the visual system chooses both alternatives in order to maintain a non-trivial yet symmetric operation: it is nonlinear and it intertwines π with $\pi \otimes \pi$. This is our reason for the discussion in Section 4.2 and for its development in the balance of this chapter.

5.2 The Mandala Transform

If we perform a Fourier transformation upon the signal output from the operator discussed in Section 4.2, then we obtain an interesting and useful result. Recall

that the operator Φ in Section 4.2 is an intertwining operator between the natural representation π of the ax+b group on the real line and its tensor square $\pi\otimes\pi$. We have that

$$\Phi(f)(x,y)=\int_{-\infty}^{\infty} k\left(\frac{y-\xi}{x-\xi}\right) f(\xi)\,\frac{d\xi}{|x-\xi|}\ .$$

Performing a Fourier transform in the second variable, we have

$$M(f)(x,\omega)=\int_{-\infty}^{\infty}\Phi(f)(x,y)\,e^{-i\omega y}\,dy=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} k\left(\frac{y-\xi}{x-\xi}\right) f(\xi)\,\frac{d\xi}{|x-\xi|}\right\}e^{-i\omega y}\,dy$$

where $M(f)$ is called the (one-dimensional) Mandala transform of f. We now proceed to manipulate the above expression into a more convenient expression. Reversing the order of the iterated integral, we obtain

$$\begin{aligned} Mf(x,\omega)&=\int_{-\infty}^{\infty} f(\xi)\int_{-\infty}^{\infty} k\left(\frac{y-\xi}{x-\xi}\right) e^{-i\omega y}\,\frac{dy}{|x-\xi|}\,d\xi \\ &=\int_{-\infty}^{\infty} f(\xi)\,e^{-i\omega\xi}\int_{-\infty}^{\infty} k\left(\frac{y-\xi}{x-\xi}\right) e^{-i\omega(x-\xi)y}\,\frac{dy}{|x-\xi|}\,d\xi\ . \end{aligned}$$

Now substituting $\frac{y-\xi}{x-\xi}$ for y with the inner integral we obtain

$$\begin{aligned} Mf(x,\omega)&=\int_{-\infty}^{\infty} f(\xi)\,e^{-i\omega\xi}\int_{-\infty}^{\infty} k(y)\,e^{-i\omega(x-\xi)y}\,dy\,d\xi \\ &=\int_{-\infty}^{\infty} h(\omega(x-\xi))\,f(\xi)\,e^{-i\omega\xi}\,d\xi \end{aligned}$$

where $h=\hat{k}$, the Fourier transform of k.

5.3 The Meaning of the Mandala Transform

The final form of the Mandala transform,

$$Mf(x,\omega)=\int_{-\infty}^{\infty} h(\omega(x-\xi))\,f(\xi)\,e^{-i\omega\xi}\,d\xi,$$

may seem familiar to many engineers. A similar expression is

$$Sf(x,\omega)=\int_{-\infty}^{\infty} h(x-\xi)\,f(\xi)\,e^{-i\omega\xi}\,d\xi$$

which is known as the Short-time Fourier Transform or the Short-time Spectrum (Portnoff 1976, Callahan 1976).

By considering the function h as a compact support "window" function in the definition of Sf, the Short-time Spectrum can be considered to be the Fourier transform of a "windowed" signal, i.e. the windowed signal is

$$f(\xi)h(x-\xi).$$

For the Mandala transform M, the process is slightly more complicated. The presence of the ω in the window means that the window changes its length for calculating each different Fourier coefficient, i.e. the windowed signal used to calculate the ω_o coefficient of the Fourier transform is

$$f(\xi)h(\omega_o(x-\xi))\equiv f(\xi)h_{\omega_o}(x-\xi).$$

Note that the window length varies in such a way so as to maintain a fixed (possibly non-integer) number of periods of $e^{i\omega\xi}$.

Vision theorists will recognize the above transform bears a strong resemblance to the receptive field model in brightness perception (Robson 1975). For a given point x_0 in space, and a given spatial frequency ω_o, the model posits a receptive field with response profile

$$h(\omega_o(x-\xi))e^{-i\omega_o\xi}$$

(or its real part), and computes the integral of the image $f(\xi)$ weighted by the response profile.

A second way of looking at the Mandala transform

follows from a simple manipulation.

$$\begin{aligned} Mf(x,\omega) &= e^{-i\omega x} e^{i\omega x} \int_{-\infty}^{\infty} f(\xi) h(\omega(x-\xi)) e^{-i\omega\xi} \, d\xi \\ &= e^{-i\omega x} \int_{-\infty}^{\infty} f(\xi) h(\omega(x-\xi)) e^{i\omega(x-\xi)} \, d\xi \\ &= e^{-i\omega x} \{ f(\xi) * [h(\omega\xi) e^{i\omega\xi}] \}. \end{aligned}$$

This last expression shows that for fixed ω, $Mf(x,\omega)$ is the baseband demodulated output of a bandpass filter with center frequency ω. Each filter has a kernel $h(\omega\xi)e^{-i\omega\xi}$ so that the bandwidth is determined by the window $h(\omega\xi)$. For larger ω, the window $h(\omega\xi)$ has a shorter impulse response, thus (by Heisenberg uncertainty (Landau and Pollack 1961)) a greater bandwidth. In fact, the bandwidth of each filter is a fixed percentage of its center frequency. For this reason, the one-dimensional Mandala transform will also be called the constant-Q transform. This second way of looking at the Mandala transform should also be familiar to vision theorists: aside from the trivial baseband demodulation, it corresponds to a special case of the filter bank model of brightness perception. Note that the discussion appearing at the beginning of this paragraph shows that the two models, receptive field and filter bank, are entirely equivalent. This fact is not commonly appreciated in the psychophysical literature (Robson 1975). Furthermore, it cannot be overstated that the Mandala transform, which bears such a close resemblance to certain psychophysical and physiological models, has been deduced solely from symmetry

considerations alone. Apart from the step of choosing the fundamental symmetry group and the a priori operator structure, the Mandala transform proceeds from purely mathematical considerations. No further experimental data was needed.

5.4 The Inverse Mandala Transform

Under rather mild restrictions on the window h, the Mandala transform is a linear operator from $W^{\infty,2}(R)$ (the Sobolev space defined in Section 4.5) into $W^{\infty,2}(R^2)$. This mapping is not onto as can be seen by Fourier transforming $Mf(x,\omega)$ with respect to x. One then obtains

$$F(\nu,\omega)=\int_{-\infty}^{\infty} Mf(x,\omega)\, e^{-i\nu x}\, dx$$
$$=F(\nu+\omega)\, k(-\frac{\nu}{\omega})$$

where $F(\cdot)$, $h(\xi)$ are the Fourier transforms of $f(\xi)$, $k(\cdot)$ respectively. Now if $h(\xi)$ is the impulse response of a lowpass filter, i.e. supp $k(\omega)\subseteq[-M,M]$ for some $M\in R$, then by the above, $F(\omega,\nu)$ is zero for $|\frac{\nu}{\omega}|>M$. This shows that Mf is not onto $L^2(R^2,\mu)$.

Because the Mandala transform is not onto $W^{\infty,2}(R^2)$, the question of an inverse (that is, a two-sided inverse) is meaningless. What we wish to show here is the existence of a <u>retract</u> (or left inverse). A retract of the Mandala transform M is a map M^{-1} such that

$$M^{-1}\circ Mf=f$$

for all $f\in W^{\infty,2}(R)$. Schematically the situation is presented in Figure 5.1.

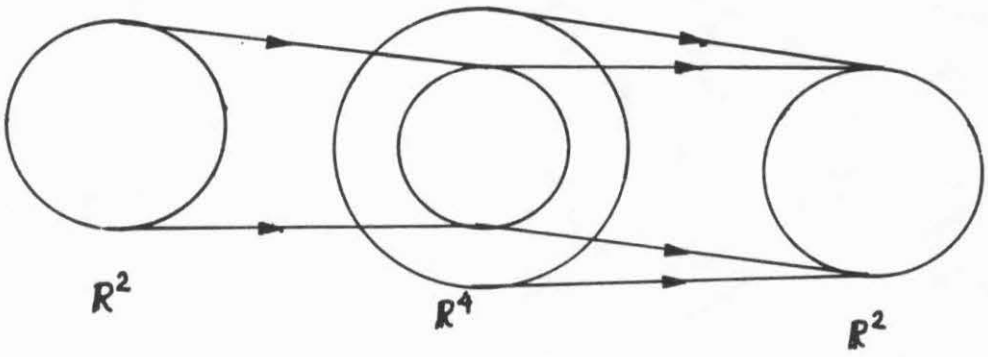


Figure 5.1 - A retraction.

In order to compute a retract of the Mandala transform, we shall first prove the following

Lemma. Let $f \in W^{\infty,2}(R)$. Then $\lim_{\epsilon \to 0} \frac{1}{-2\log\epsilon} \int_{\epsilon<|\eta|} \frac{f(\eta)}{|\eta|} d\eta = f(0)$.

Proof. We shall separate the integral into two parts and then compute each part.

$$(5.4.1) \qquad \int_{\epsilon<|\eta|} \frac{f(\eta)}{|\eta|} d\eta = \int_{\epsilon<|\eta|<1} \frac{f(\eta)}{|\eta|} d\eta + \int_{1\le|\eta|<\infty} \frac{f(\eta)}{|\eta|} d\eta.$$

To estimate the last term of the RHS of (5.4.1), notice that $f(\eta), \frac{1}{|\eta|} \in L^2([1,\infty])$, thus we can use the Hölder inequality

$$\int_{1\le|\eta|<\infty} \frac{f(\eta)}{|\eta|} d\eta \le \|f\|_{[1,\infty)} \cdot \|\tfrac{1}{|\eta|}\|_{[1,\infty)} + \|f\|_{(-\infty,1]} \cdot \|\tfrac{1}{|\eta|}\|_{(-\infty,1]} \le K_1 < \infty.$$

For the first term we have by Taylor's theorem

$$\begin{aligned} \int_{\epsilon<|\eta|<1} \frac{f(\eta)}{|\eta|} d\eta &= \int_{\epsilon<|\eta|<1} \frac{f(0)+\eta f'(0)+R(\eta)}{|\eta|} d\eta \\ &\le f(0)(-2\log\epsilon) + \sup_{(0,1)} f''(x) + \sup_{(-1,0)} f''(x) \\ &= f(0)(-2\log\epsilon) + K_2. \end{aligned}$$

The inequality follows from Taylor's theorem and the mean value theorem. Thus we have that

$$\begin{aligned} &\lim_{\epsilon\to 0} \frac{1}{-2\log\epsilon} \int_{\epsilon<|\eta|} \frac{f(\eta)}{|\eta|} d\eta \\ &\lim_{\epsilon\to 0} \frac{1}{-2\log\epsilon} (f(0)\cdot(-2\log\epsilon) + K_1 + K_2) \\ &= f(0). \end{aligned}$$

Replacing $f(\eta)$ with $-f(\eta)$ in the previous steps gives the lemma.

Note that the lemma says in effect that the function $\delta_\epsilon(\eta)$ defined by

$$\delta_\epsilon(\eta)=\begin{cases}\dfrac{1}{-2|\eta|\log\epsilon} & \epsilon<|\eta| \\ 0 & \epsilon>|\eta|\end{cases}$$

is an approximate identity.

From the symmetry theory we have that all retracts with the a priori structure

$$\int G(\xi,x,\omega)\,Mf(x,\omega)\,dxd\omega$$

have the kernel

$$G(\xi,x,\omega)=g(\omega(x-\xi))\,e^{i\omega\xi}\,|\omega|\ .$$

In fact, these kernels do give retracts as the following proposition shows.

Proposition: The transform

$$M^{-1}f(\xi)=\lim_{\epsilon\to 0}\frac{1}{K\log\epsilon}\iint_{\epsilon<\frac{|\omega(x-\xi)|-M}{|\omega|}} g(\omega(x-\xi))\,e^{-i\omega\xi}\,|\omega|\,f(x,\omega)\,dxd\omega$$

is a retract of the Mandala transform, where

$$K=-2\left[\int g(s)\,e^{-is}\,ds\right]\left[\int h(s)\,e^{is}\,ds\right]=-2k(1)\left[\int g(s)\,e^{-is}\,ds\right]$$

and $[-M,M]$ is the support of $h(\xi)$.

Proof. We substitute the definition of the Mandala transform and use freshman calculus, evaluating the integral

$$\iint_{\epsilon<\frac{|\omega(x-\xi)|}{|\omega|}} g(\omega(x-\xi))\,e^{i\omega\xi}\,|\omega|\int_{-\infty}^{\infty} f(\sigma)\,h(\omega(x-\sigma))\,e^{-i\omega\xi}\,d\sigma dxd\omega\ .$$

Since for $|\omega(x-\sigma)|>M$ the inner integrand is zero, the actual limits of integration are $|\omega(x-\sigma)|<M$.

Now we exchange integrals

$$= \int_A f(\sigma) \iint_B g(\omega(x-\xi)) h(\omega(x-\sigma)) e^{i\omega(\xi-\sigma)} \, dx d\omega d\sigma$$

where the sets A,B have yet to be determined.

Making the substitution

$$u = \omega(x-\xi)$$

$$v = \omega(x-\sigma)$$

we get $|J| = \left| \frac{\partial(x,\omega)}{\partial(u,v)} \right| = \frac{1}{|\omega| \, |\xi-\sigma|}$.

Thus the integral becomes

$$\int_A f(\sigma) \iint_B g(u) h(v) e^{i(v-u)} \frac{du \, dv}{|\xi-\sigma|} \, d\sigma .$$

To determine A,B, we start with the limits given at the outset

$$\epsilon < \frac{|\omega(x-\xi)| - M}{|\omega(x-\sigma)| < M.}$$

Thus

$$\epsilon < \frac{|\omega(x-\xi)| - |\omega(x-\sigma)|}{|\omega|} = |x-\xi| - |x-\sigma| \leq |\sigma-\xi| = |\xi-\sigma|$$

$$|u| = |\omega(x-\xi)| > M + |\omega|\epsilon = M + \left|\frac{v-u}{\xi-\sigma}\right| \epsilon$$

$$|v| < M.$$

or in other words

$$A = \{\sigma \mid |\xi-\sigma| > \epsilon\}$$

$$B = \{(u,v) \mid |v| < M; \; M + \left|\frac{v-u}{\xi-\sigma}\right| \epsilon < |u|\} .$$

The set B appears in Figure 5.2. We may now rewrite the integral as

$$\int_{\epsilon < |\xi-\sigma|} \frac{f(\sigma)}{|\xi-\sigma|} \left[\int_{-M}^{M} \int_{-M}^{M} g(u) h(v) e^{-iu} \, e^{iv} \, du dv \right.$$

$$\left. - \int_{-M}^{M} \int_{C} g(u) h(v) e^{-iu} \, e^{iv} \, du dv \right] d\sigma$$

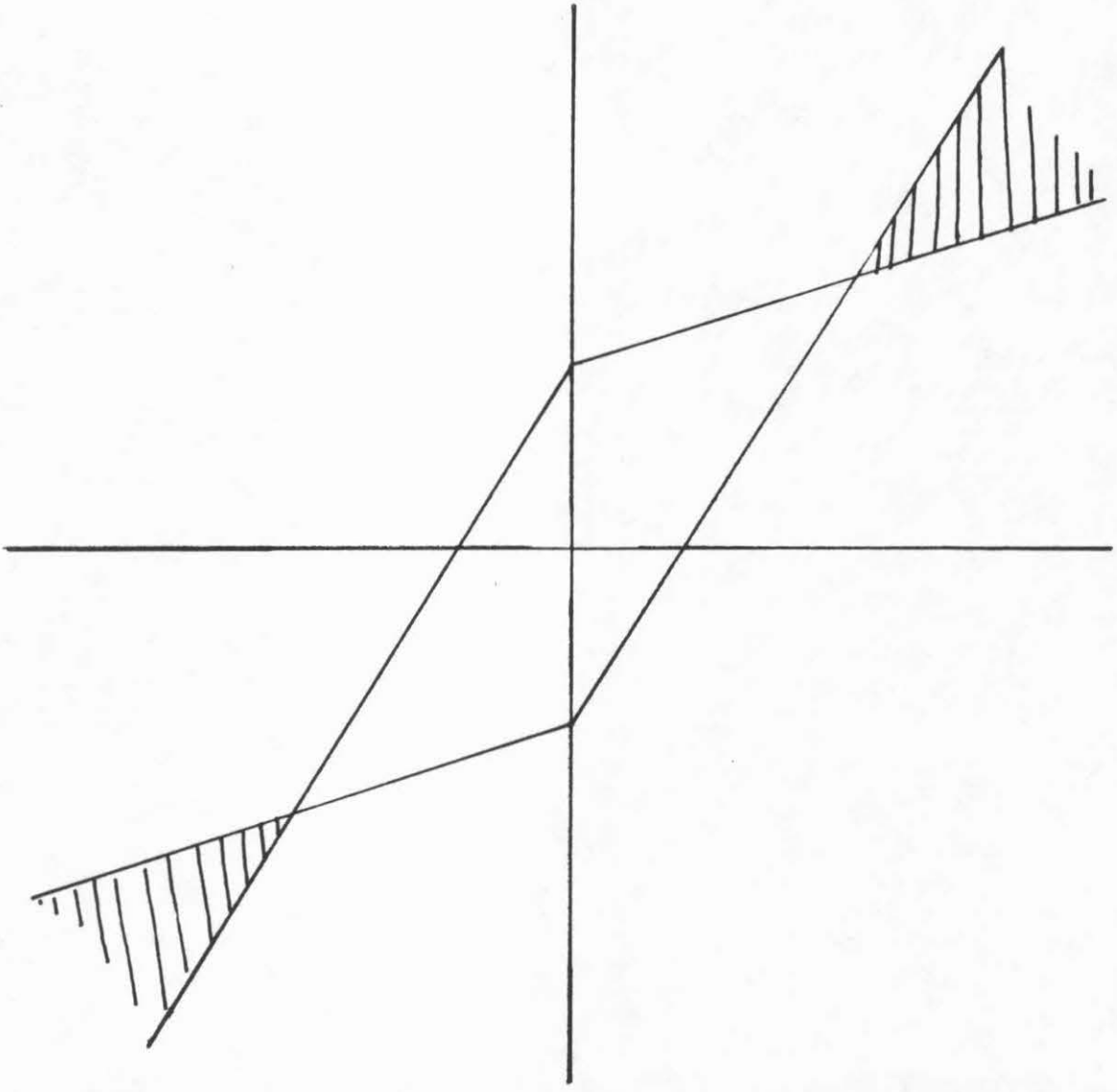


Figure 5.2 - Integration regions.

$$= \int_{\epsilon<|\xi-\sigma|} \frac{f(\sigma)}{|\xi-\sigma|}\, d\sigma \quad g(u)\, e^{-iu}\, du \int_C h(v)\, e^{iv}\, dv$$

$$- \int_{\epsilon<|\xi-\sigma|} \frac{f(\sigma)}{|\xi-\sigma|} \int_{-M}^{M}\int_C g(u)\, h(v)\, e^{-iu}\, e^{iv}\, du\, dv\, d\sigma$$

where $C=\{(u,v)|\ |v|<M;\ M<|u|<M+\frac{|v-u|}{|\xi-\sigma|}\epsilon\}$.

The last term in the above equation can be estimated as less than

$$-2M^2\epsilon \sup_C |g(u)\, h(v)| \int_{\epsilon<|\xi-\sigma|} \frac{f(\sigma)}{|\xi-\sigma|^2}\, d\sigma$$

$$\leq -2M^2 \sup_C |g(u)\, h(v)| \sup_A |f(\sigma)|\epsilon \int_{\epsilon<|\xi-\sigma|} \frac{1}{|\xi-\sigma|^2}\, d\sigma$$

$$= +2M \sup_C |g(u)\, h(v)| \sup_A f(\sigma) < \infty.$$

The proposition now follows by dividing by $K \log\epsilon$: the second term estimated above tends to zero, the first term gives the correct result when applying the lemma.

5.5 The Windowing Approximation

This section will discuss the relationship between the Short-time Fourier transform and the Mandala transform.

An important motivation for the Short-time Fourier transform stems from the desire to compute the Fourier transform in an effective manner. That is, how can one calculate an approximation to

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(\xi)\, e^{-i\omega\xi}\, d\xi$$

on a digital computer? The first obvious approximation is to convert the integral to a sum

$$\hat{f}(\omega) \sim \sum_{n=-\infty}^{\infty} f(nT)\, e^{-i\omega nT}$$

where T is some suitably chosen "sampling period". Even with this simplification an infinite sum remains, therefore in order to make the calculation in a reasonable amount of time, the sum is truncated to a finite number of terms:

$$\hat{f}(\omega) \sim \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} f(nT)\, e^{-i\omega nT} .$$

This second approximation is equivalent to truncating the input signal to a finite record length, or, in other words, windowing the input signal with a Fourier window

$$h_{fourier}(\xi) = \begin{cases} 1 & |\xi| \leq \frac{N}{2} \\ 0 & |\xi| > \frac{N}{2}. \end{cases}$$

Thus the formula now reads

$$\hat{f}(\omega) = \sum_{n=-\infty}^{\infty} f(nT)\, h_{fourier}(nT)\, e^{-i\omega nT} .$$

To center the window about a point other than zero, we translate the window

$$\hat{f}(x,\omega) = \sum_{n=-\infty}^{\infty} f(nT)\, h_{fourier}(x-nT)\, e^{-i\omega nT} .$$

The reader will easily recognize this as the sampled version of the Short-time Fourier transform.

We are now confronted with a problem: How do we choose our two approximations wisely in order to maintain accuracy in our calculations? For the first approximation, sampling the input signal, the celebrated Shannon sampling theorem provides a reliable guide (Shannon 1949). The second approximation is somewhat more problematical. Choosing a given function as a window involves many

considerations (Blackman and Tukey 1968), but the only aspect with which we shall be concerned is the window length. There is no analog of the Shannon sampling theorem to help us decide how long to make the window for any given application. Indeed, in most applications it is apparent that a single optimum window length does not exist, i.e. a single given window length will be simultaneously too short in some respects and too long in others.

The key objection to the windowing approximation is, however, given by the central theme of this thesis -- symmetry. It is easily seen that the (unapproximated) Fourier transform possesses symmetry under two representations of the ax+b group. Thus for both perceptual modelling and processing of signals, the Fourier transform is important. The fundamental objection to the windowing approximation is that it destroys the dilatation symmetry enjoyed by the Fourier transform. In fact, the symmetry operations corresponding to the Short-time Fourier transform are translation and multiplication by $e^{i\alpha\xi}$, i.e. complex modulation.

Since the Mandala transform was designed to be symmetric, we believe it to be the proper approximation to the Fourier transform for computer purposes. Furthermore, the window length now changes for each analysis frequency, becoming shorter for higher frequencies and longer for

lower frequencies. It turns out that, for applications, this is precisely the behavior that is needed.

CHAPTER 6

THE HUMAN VISUAL SYSTEM
BRIGHTNESS PERCEPTION PHENOMENA AND MODELS

6.1 Introduction

It is well known that the judgment of human beings concerning the absolute intensity of light falling upon the retina generated by some two-dimensional distribution of light can be very inaccurate. We have all amused ourselves at one time or another by various brightness illusions. For example, a given spot of light of fixed intensity may be judged bright or dim depending upon its surroundings. These brightness illusions show that the exact correspondence between the perceived intensity or brightness of a given stimulus and the actual physical intensity or luminance is not as simple as the naive observer may assume. This correspondence is the subject of this and subsequent chapters. In particular, this chapter will present some established phenomena and new phenomena in brightness perception. Also, several popular models of brightness perception will be presented and their performance with respect to the above phenomena will be discussed.

Section 6.2 will present various well-known brightness illusions, that is, images which are particularly striking in that their brightness distributions are considerably different from the actual luminance distributions. These images alert us to the fact that there is a mechanism which is worthy of study. The next section, 6.3, will describe a set of experiments which have been particularly important in the last decade or so. These experiments measure the ability of the human visual system to detect the presence of small changes of some stimulus (detection experiments) and how that ability is altered by an initial exposure of various stimuli (adaptation experiments). These studies are extremely important because their results are highly repeatable. The next section will present new phenomena in brightness perception discovered in the course of the present research. Finally, the last section will discuss several models or, more accurately, classes of models and evaluate their performance. That is, a given model may be rated as excellent or poor according to the similarity of its response with respect to the visual system.

6.2 Illusions

We will discuss a number of intuitively interrelated illusions, all displaying some aspect of the human visual system.

Mach Bands. The illusion illustrated in Figure 6.1 is the celebrated Mach band illusion (Ratliff 1965). The Mach bands themselves appear as accentuated bright or dark bands near the C^1-discontinuities of the intensity distribution profiled directly below the image. As Mach himself points out (Mach 1886), this sensitivity to C^1-discontinuities and relative insensitivity to continuous changes in brightness serves to facilitate the delimitation of objects by enhancing their "contours". These facts led Mach to the conclusion that the response of each retinal element is not independent of the responses of its neighbors but is dependent through a "reciprocal action". These observations form the basis of the lateral inhibition model discussed in Section 5.

A significant feature of Mach bands is that they disappear (or at least become very narrow) at an "edge", i.e. a C^0-discontinuity in the image (Davidson and Whiteside 1971). Figure 6.2 shows the disappearance of the Mach band at an edge. Figure 6.3 interpolates the previous two images so that one may follow the genesis of the Mach band.

Edge Scalloping. The image in Figure 6.4 gives rise to an illusion that is related to but not identical with the Mach band illusion. As the reader can easily observe, the areas of uniform luminance as indicated by the image

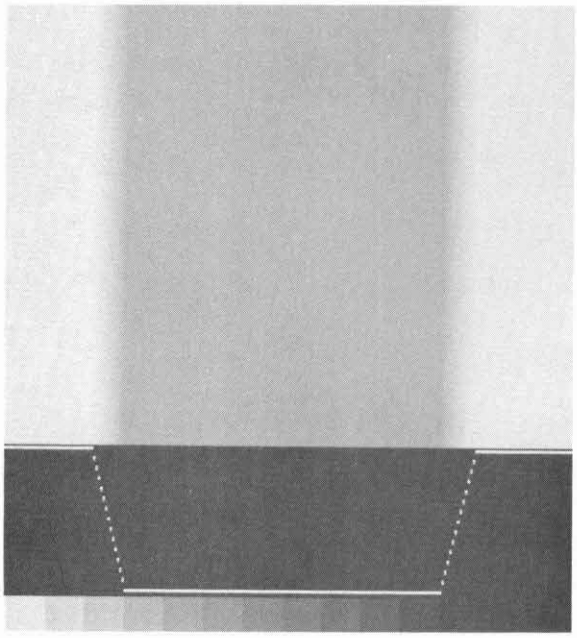


Figure 6.1 - Mach bands.

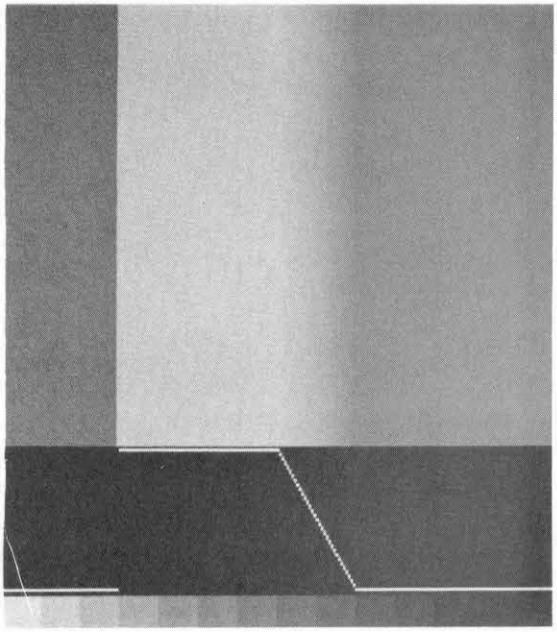


Figure 6.2 - Mach band with edge.

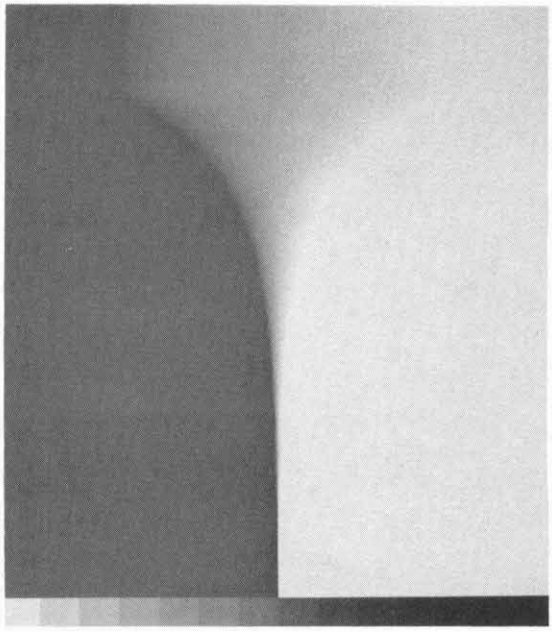


Figure - 6.3 - Transition from ramp to edge.

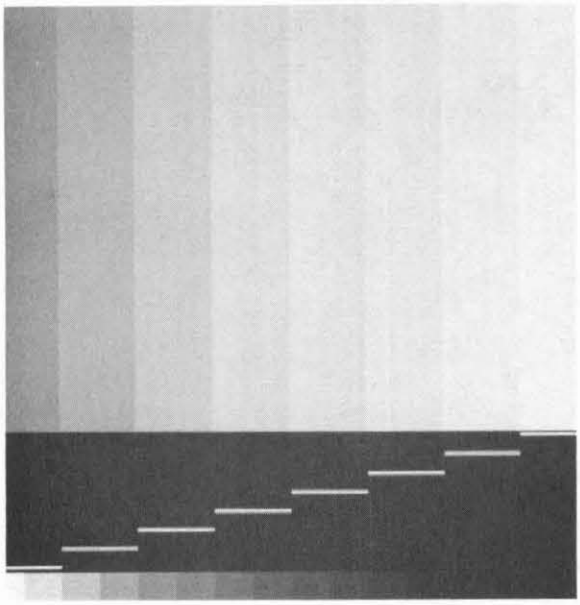


Figure 6.4 - Edge scalloping.

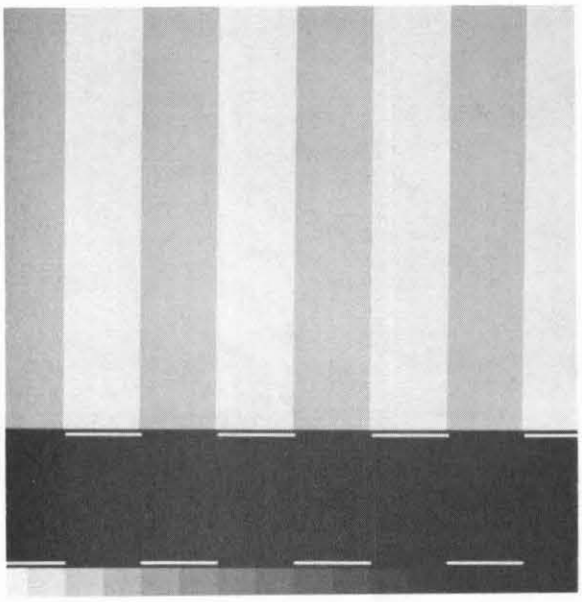


Figure 6.5 - Square wave train.

profile induce a nonuniform brightness response. Furthermore, as pointed out by Bekesy (1968), if all but a single edge is covered up by, say, a white card, the brightness response becomes uniform. As successive edges are uncovered, the areas of uniform luminance appear more and more scalloped.

This effect is not shared by an image in which the same number of edges is present, but the values of the jumps alternate in sign (see Figure 6.5).

Cornsweet Illusion. This illusion, which is also associated with the names Craik, Clark and O'Brien (Cornsweet 1970, Ratliff 1971), is displayed in Figures 6.6 and 6.7. Briefly, the luminance of the image is uniform at a distance from the contours, but the perceived brightness of the image is uniform at some much lesser distance from the contours. Furthermore, the center of the image appears darker than the periphery, yet the luminance of the two areas is exactly the same, as can be seen by the reader simply by constructing a mask covering the contour. This illusion is particularly striking when one considers that a purely "local" stimulus (the contour) is causing a "global" illusion (the disparity in brightness between center and periphery).

An illusion closely related to Cornsweet's is the missing fundamental illusion (Campbell, Howell and Robson

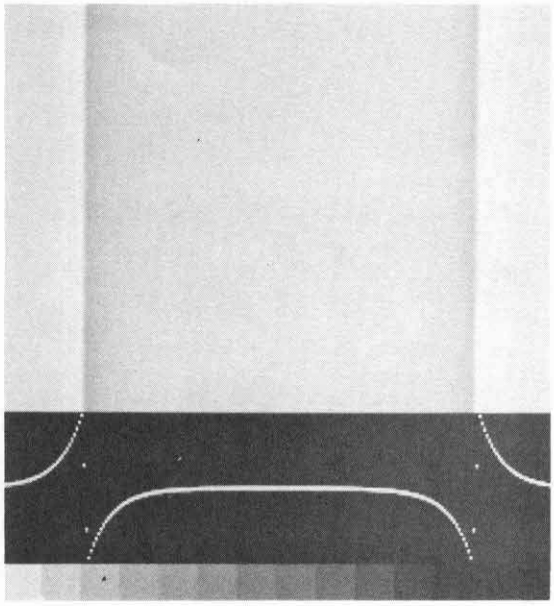


Figure 6.6 - Cornsweet illusion.

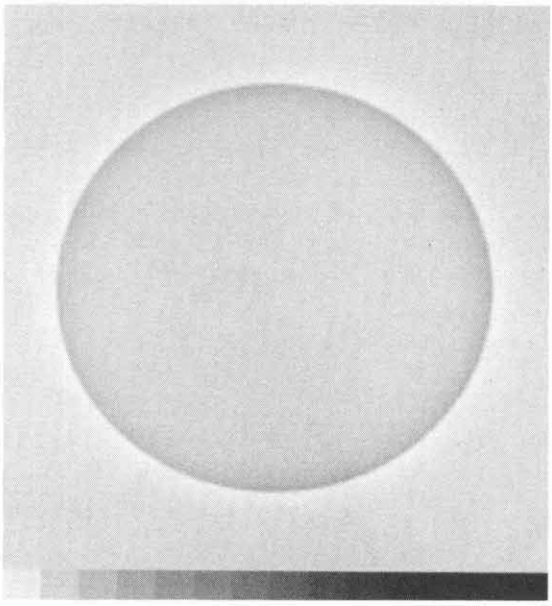


Figure 6.7 - Cornsweet illusion.

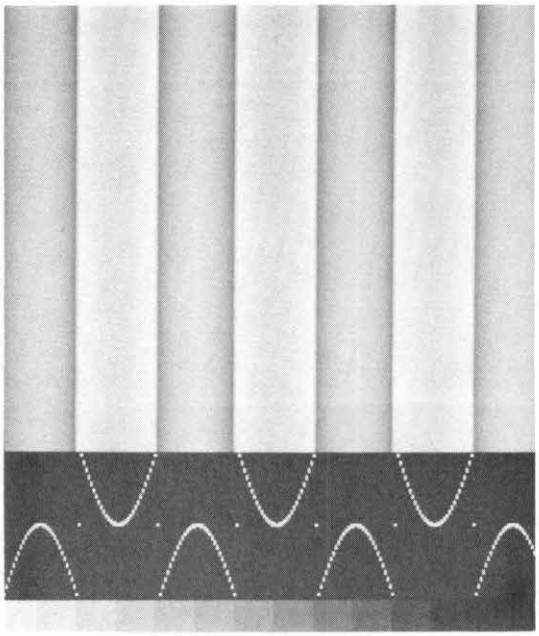


Figure 6.8 - Missing fundamental illusion.

1971). This illusion is displayed in Figure 6.8. The image is created simply by summing the Fourier series for a square wave but leaving out the first term, i.e. the fundamental. As the image profile shows, the areas between the edges are heavily scalloped yet the perceived brightness does not possess this scalloping.

6.3 Detection/Adaptation Experiments

A particularly important mechanism of the visual system is the thresholding of signals within the visual system. If two stimuli S_1, S_2 are presented to the visual system, one identical with the other except for the addition of a component of small amplitude, then a detection experiment seeks to determine at what amplitude the additional component must be in order for the observer to be able to discern the difference between these two stimuli. This is usually accomplished by presenting the two stimuli in succession and requiring the subject to respond yes or no according to his perception of a difference between the stimuli. The above sequence, or trial, is repeated many times, each with a random setting of the amplitude of the additional component (including, of course, zero amplitude, which results in no difference between the two stimuli). The results of each trial at each particular amplitude are then tabulated. The 50% correct amplitude is called the threshold of that component.

The second experiment attempts to measure any shift in the threshold for a particular component by first adapting the visual system to another component. Thus, this experiment modifies the above trial by requiring the observer to view a prolonged adapting stimulus S_a before performing the procedure described above. Results are then tabulated and the threshold shift (which may be zero) is calculated.

Detection experiments. Starting with DePalma and Lowry (1962), many experimenters have performed the detection experiment with stimuli

$$S_1 : L_0$$

$$S_2 : L_0 (1+m\sin\omega x)$$

where L_0 is the mean luminance, while m is the contrast, and ω is the spatial frequency of the additional component known as a grating.

For a given L_0, ω_0 , the contrast threshold m is measured by the detection experiment. Figure 6.9 shows the result of such an experiment, where instead of the contrast threshold m, the contrast sensitivity $\frac{1}{m}$ is displayed. These curves are taken from Campbell and Robson (1968). It is interesting that contrast sensitivity (thus contrast threshold) is strongly affected by spatial frequency but is nearly independent of mean luminance (Van Nes and Bowman 1965).

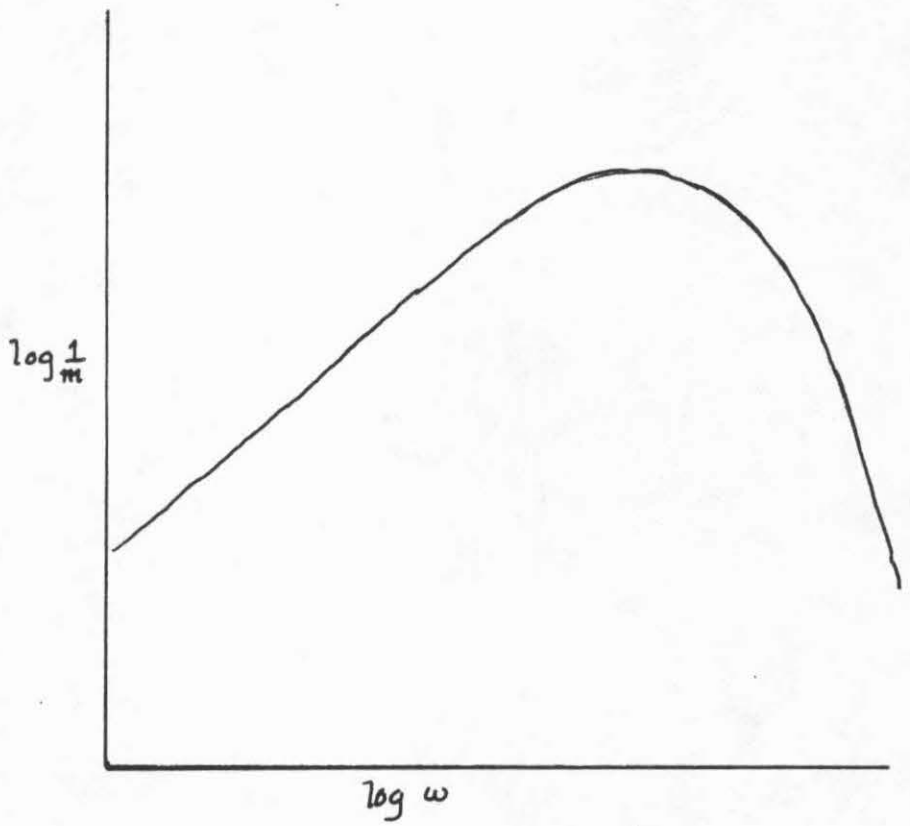


Figure 6.9 - Contrast sensitivity of the visual system (after Campbell and Robson 1968)

A significant variation of the above experiment was performed by Sachs, Nachmias and Robson (1971). The stimuli in this detection experiment differs by forming the additional component from not a single sinusoid but two sinusoids.

$$S_1 : L_0$$

$$S_2 : L_0(1+m\sin\omega x+m'\sin(\omega' x+\phi))$$

where L_0 is the mean luminance, m and m' are the contrasts, ω and ω' are the spatial frequencies, and ϕ is the relative phase of the gratings. The key result of the above experiment is that, for sufficient separation in spatial frequencies ω, ω', the detectability of the additional component is determined by the detectability of each grating independently. That is, if either m or m' is at threshold for detection in the single grating case, then the component is at threshold provided that the separation between ω and ω' is sufficiently great. This separation must be approximately a ratio $\frac{\omega'}{\omega}<\frac{1}{2}$ or $\frac{\omega'}{\omega}>2$. Furthermore, when the separation between ω and ω' is less than this ratio, then the contrasts m and m' combine in such a way so as to determine the threshold of the complex grating, i.e. am+a'm' determines the detectability of the complex grating where a and a' depend upon ω and ω'.

Adaptation Experiments. In a landmark paper, Blakemore and Campbell (1969) report on an adaptation experiment performed with the following stimuli:

$$S_a = L_o (1 + m_a \sin \omega_a x)$$

$$S_1 = L_o$$

$$S_2 = L_o (1 + m_t \sin \omega_t x)$$

where L_o is the mean luminance, m_a and m_t are the adapting contrast and test contrast respectively, ω_a and ω_t are the adapting spatial frequency and test spatial frequency. The increase in contrast threshold is calculated as

$$\Delta = \log \frac{\bar{m}_t'}{\bar{m}_t}$$

where $\bar{m}_t$ is the threshold contrast without adaptation and $\bar{m}_t'$ is the threshold contrast after adaptation. Blakemore and Campbell fixed L_o, m_a, ω_a and measured Δ for various ω_t. A typical curve is redrawn in Figure 6.10. As can be clearly seen, an adapting grating only raises the threshold for test gratings of similar spatial frequencies. Furthermore, for different adapting frequencies ω_a, the contrast shift curves plotted against various test frequencies show surprising regularity (see Figure 6.11). Finally, as the adaptation contrast is varied, the threshold shift curve varies. In Figure 6.12, several threshold shift curves corresponding to various adaptation contrasts are shown. Note that when their peaks are shifted, they show a remarkable shape similarity.

Blakemore and Campbell also performed the adaptation experiment for a square wave grating and found that there was a corresponding threshold rise in the first and third

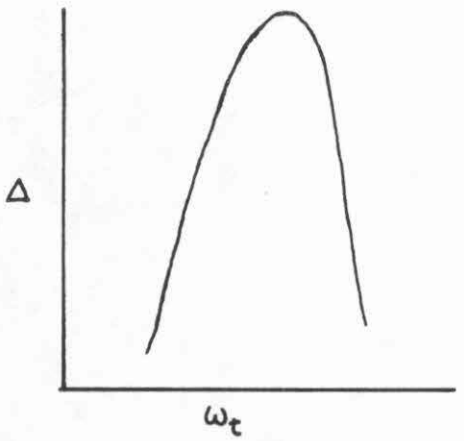


Figure 6.10 - Adaptation response vs. test frequency (after Blakemore and Campbell 1969)

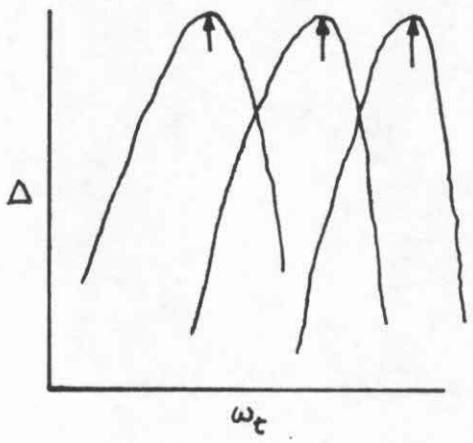


Figure 6.11 - Adaptation response vs. adaptation frequency

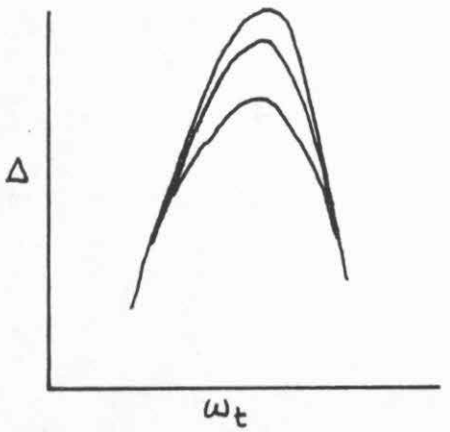


Figure 6.12 - Adaptation response vs. adapting contrast

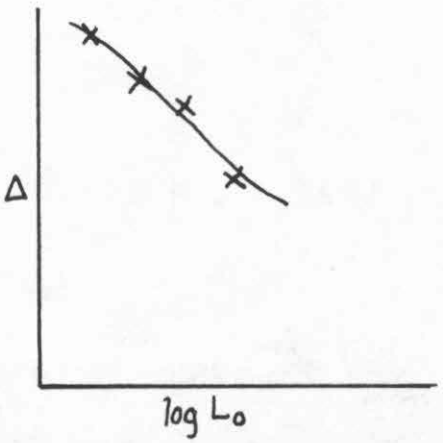


Figure 6.13 - Adaptation response vs. mean luminance

harmonic test spatial frequencies. Tolhurst (1972) explores this situation further by attempting to measure the third harmonic contrast threshold shift and relating it to the contrast of the third harmonic "contained" in the square wave (calculated by a Fourier series development). Tolhurst found that the third harmonic (as well as the fundamental) was much less effective as an adapting stimulus when present in a square wave stimulus than when presented alone in a simple sinusoidal grating. Tolhurst then goes on to construct an adaptation experiment using simpler (i.e. with less terms in the Fourier series) adapting stimuli than a square wave:

$$S_a : L_o\left(1+\frac{4m_a}{\pi}\sin\omega_a x+\frac{4m_a}{3\pi}\sin 3\omega_a x\right)$$

$$S_1 : L_o$$

$$S_2 : L_o(1+m_t \sin\omega_t x)$$

Tolhurst explored the case for a fixed L_o, and $\omega_t=\omega_a$ or $\omega_t=3\omega_a$. Indeed, as the square wave case would suggest, the adapting sinusoids were much less effective in the complex grating than when they were presented in a single grating.

Both Blakemore and Campbell (1969) and Tolhurst (1972) performed their adaptation experiments with a fixed mean luminance L_o of 100 cd/m^2. A natural question arises: How does the adaptation effect vary with mean luminance? Maudarbocus and Ruddock (1973) attempt to measure this dependence. A typical curve is shown in

Figure 6.13. In it, the adapting contrast is held fixed and $\omega_a = \omega_t$ while L_o varies. Thus the effect of increasing the mean luminance is to decrease the adaptation effect. Furthermore, Maudarbocus and Ruddock found that for extreme values of adapting contrast, the frequency selectivity of the threshold adaptation curve did not change. This fact will be important when discussing possible models.

Legge (1976) performed the adaptation experiment with an adapting stimulus of a single narrow bar of high luminance and test stimuli sinusoidal gratings:

$$S_a : L_o + \delta(x)$$

$$S_1 : L_o$$

$$S_2 : L_o(1+m\sin\omega x)$$

where $\delta(x)$ is a bright bar of width 0.6' and luminance 140 cd/m^2, L_o is the mean luminance (4.63 cd/m^2) and the spatial frequency of the test grating ω is set to various values. Legge found that adaptation to a narrow bar resulted in a very slight, almost insignificant, threshold shift at all test frequencies.

6.4 New Phenomena

This section will present a number of new phenomena which are either interesting in themselves or important for the discussion on models in the following section.

Illusion Softening. A procedure which increases our

understanding of the mechanism responsible for a given illusion is one we have called illusion softening. The procedure is simply to surround a given illusion with various stimuli in order to lessen or cancel the illusion to obtain a brightness perception similar to the actual luminance distribution. The first example of illusion softening is shown in Figures 6.14 and 6.15. This example is not new. It is a well-known effect of the Cornsweet illusion. The image displayed is a series of Cornsweet edges. When one observes this image, the perceived brightness of the image is not a stepwedge, as one would expect from a series of Cornsweet edges, but simply a uniform background with a series of scalloped edges superimposed. If, however, all but one edge is covered up, then the familiar disparity in brightness on either side of the remaining uncovered edge occurs. Thus the underlying mechanism governing the Cornsweet illusion is dependent not only on purely local effects but also on the global distribution of light intensity throughout the image. Note that if the Cornsweet edges alternate in sign, then illusion softening does not occur (see Figure 6.16). It is not only additional Cornsweet edges which soften the Cornsweet illusion. One may also surround a Cornsweet edge by actual luminance steps and obtain the softening effect (see Figure 6.17).

Not only the Cornsweet illusion is subject to

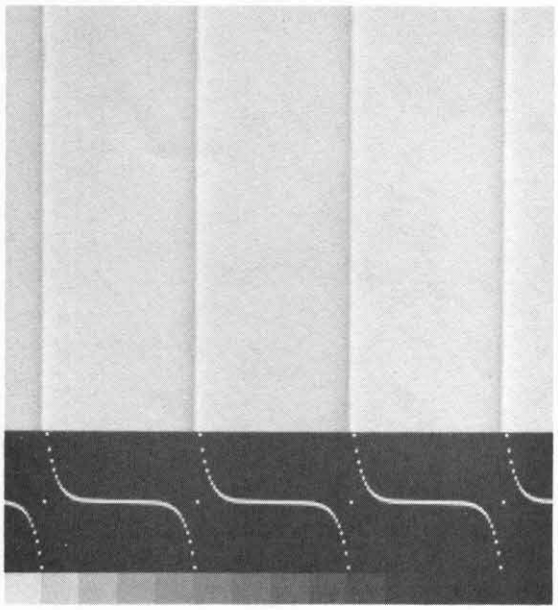


Figure 6.14 - Illusion softening: the Cornsweet illusion.

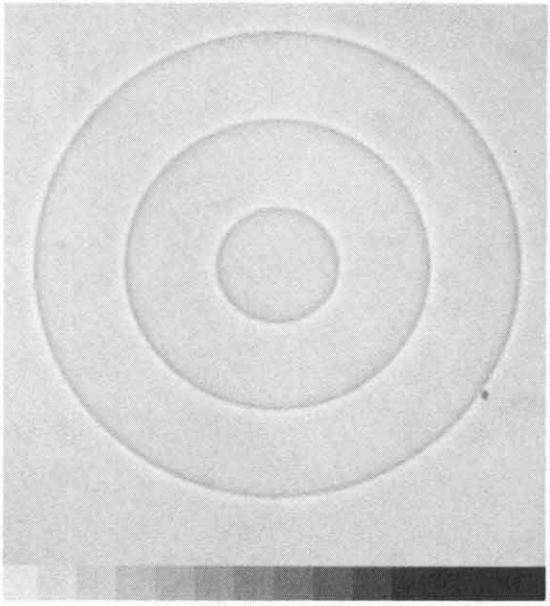


Figure 6.15 - Illusion softening: the Cornsweet illusion.

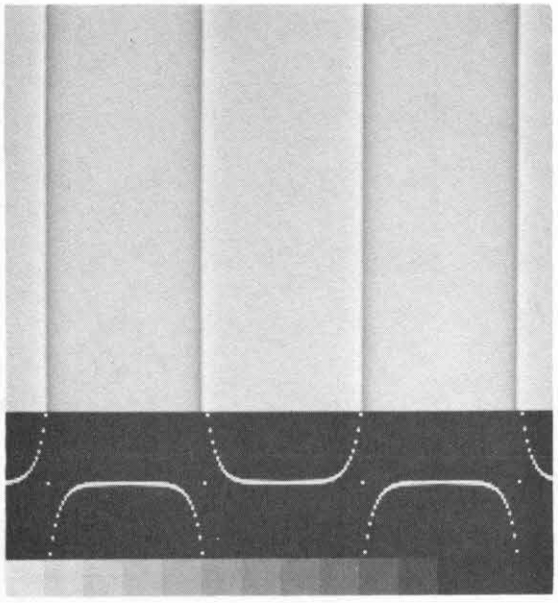


Figure 6.16 - No illusion softening: the Cornsweet illusion.

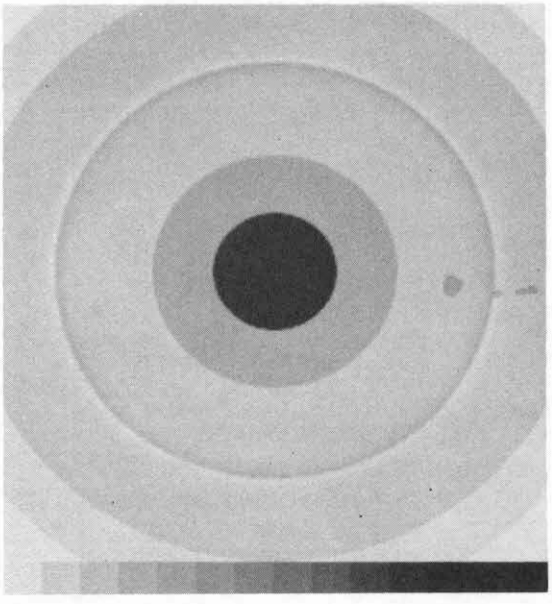


Figure 6.17 - Illusion softening: the Cornsweet illusion.

softening. The oldest and most studied brightness illusion also can be softened. That the Mach band illusion can be softened is a new result of this research and not a well-known classical result is somewhat puzzling given its status as the most intensively studied illusion since its discovery over one hundred years ago. A typical Mach band illusion is shown in Figure 6.18 (where only a section of 6.1 is shown). The softening of the dark Mach bands can be accomplished by surrounding Figure 6.18 by a luminance step; this image is shown in Figure 6.19. Note that the dark Mach bands have been softened in this image but the white Mach bands remain intact. Figure 6.20 shows both the white and dark Mach bands softened. The center strip now has a perceived brightness remarkably similar to the actual luminance profile of the original image (as in Figure 6.17). The reader may amuse himself by covering up the surrounding stimuli to reestablish the Mach illusion (which, interestingly enough, requires an amount of time to reinstate itself).

New Illusion. A new illusion is shown in Figures 6.21 and 6.22. In 6.22, the luminance profile is shown, which is also the luminance profile along any diameter of 6.21. The illusion is very similar to the Cornsweet and Missing Fundamental illusions except that no sharp luminance discontinuity is present in the image. Despite this, the

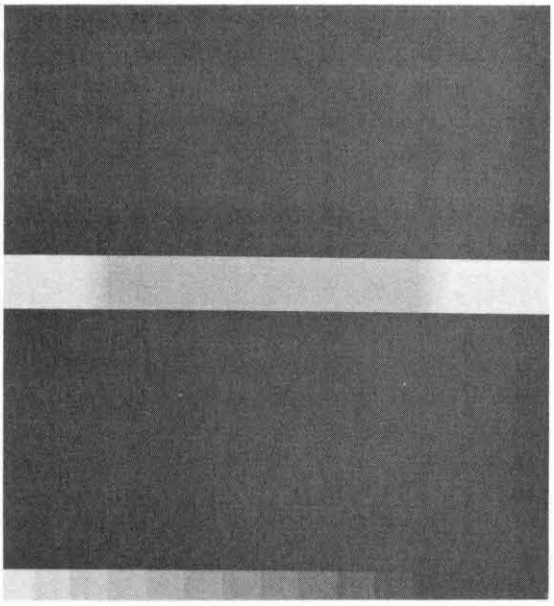


Figure 6.18 - Mach bands.

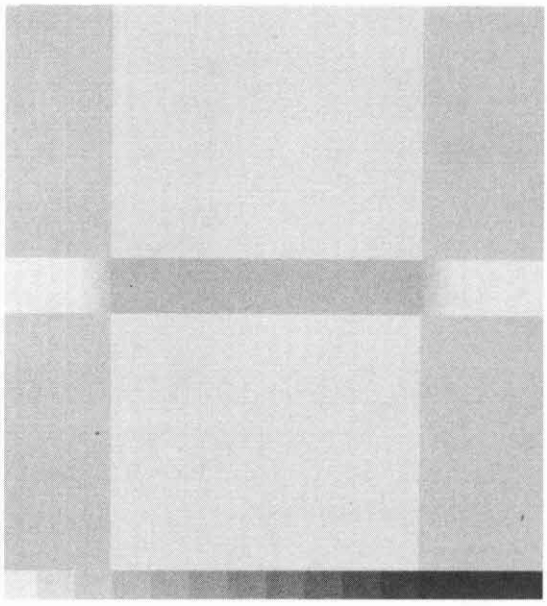


Figure 6.19 - Illusion softening: the dark Mach band.

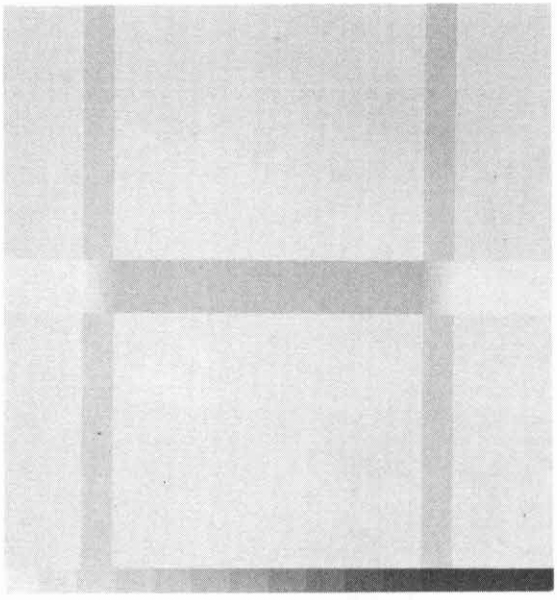


Figure 6.20 - Illusion softening: both Mach bands.

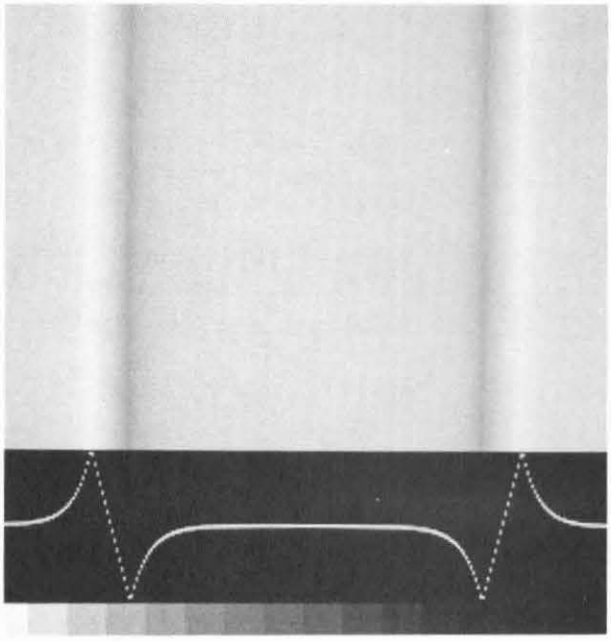


Figure 6.21 - The new illusion.

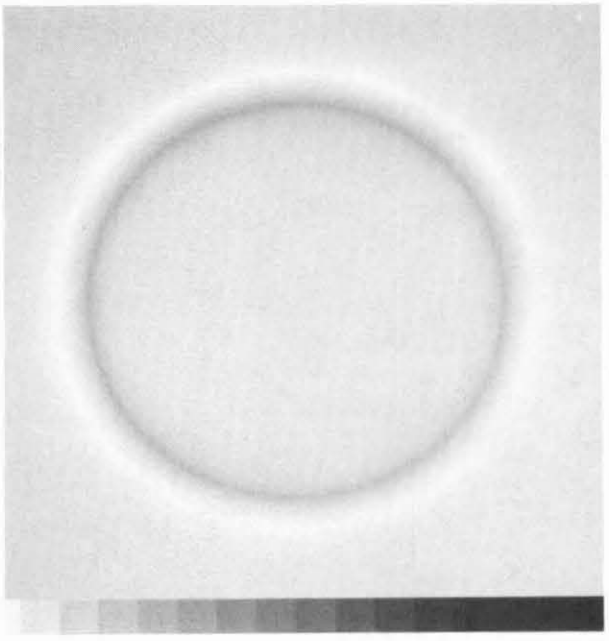


Figure 6.22 - The new illusion.

perceived brightness of the center of the image is darker than the periphery. Covering up the non-uniform parts of the image quickly brings the perceived brightness in line with the actual luminance, viz. the center and periphery appear the same brightness.

Adaptation Phenomena. Aside from adaptation to square-wave and sinusoidal gratings and impulses described above, the only other adaptation experiments of which the author is aware have been performed by Weisstein and Bisaha (1972), Sullivan, et al.(1972) in which single bars have been used as adapting stimuli. During the course of the present research, the author has become convinced that an adaptation experiment involving adapting stimuli as a rectangular series of pulses with various fixed duty cycles will yield important data. In lieu of a complete psychophysical experiment (which evidently requires rather elaborate psychophysical apparatus and technique), we may perform the rather crude experiment described below.

Figure 6.23 shows the familiar image of Campbell and Robson's that displays the modulation transfer function of the visual system. The image profile at any given scan line is that of a chirp signal, i.e. a linear FM grating. For each line the grating maintains a fixed contrast. Along the downward vertical direction, the contrast decreases in a logarithmic manner. Thus the curve joining

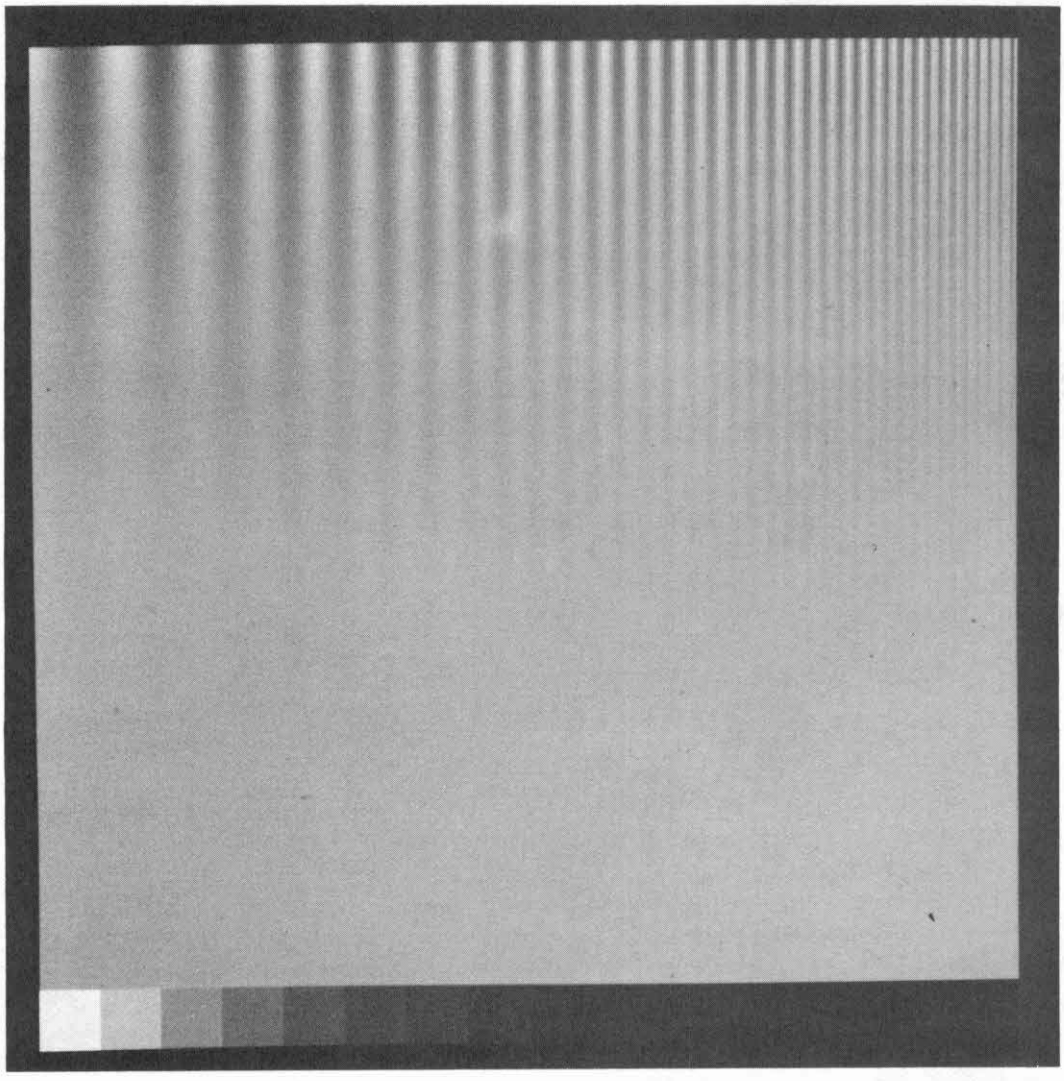


Figure 6.23 - Contrast threshold vs. frequency.

the apparent disappearances of each bar will trace out the threshold for detection for that spatial frequency. We shall call this curve, for convenience, the thresholding curve. This figure is meant to be viewed at a distance of 50 cm. With the aid of this figure, we may now conduct a crude adaptation experiment:

1) View the image displayed in Figure 6.23 at the proper distance and note the position of the thresholding curve.

2) View the grating displayed in Figure 6.24 for a period of approximately 50-100 seconds.

3) The thresholding curve will no longer appear flat but have a bulge at approximately the 3 cycle/degree abscissa. The difference between the new thresholding curve and the old thresholding curve gives the threshold rise (this holds, of course, even if the original thresholding curve is not flat). The reader has, if he has performed the above steps, just conducted a crude approximation to one of the classical adaptation experiments with a sinusoidal grating adapting stimulus. Now the adaptation experiment with adapting stimuli as rectangular pulses with various duty cycles can be performed. We will be primarily interested in the extreme case of almost 100% duty cycle. The reader may perform this experiment by substituting Figure 6.25 in Step 2.

The entire experiment may be condensed in a way by resorting to a masking paradigm. Various sine wave

Figure 6.24 - 3 cy/deg sinusoidal grating (at 50 cm).

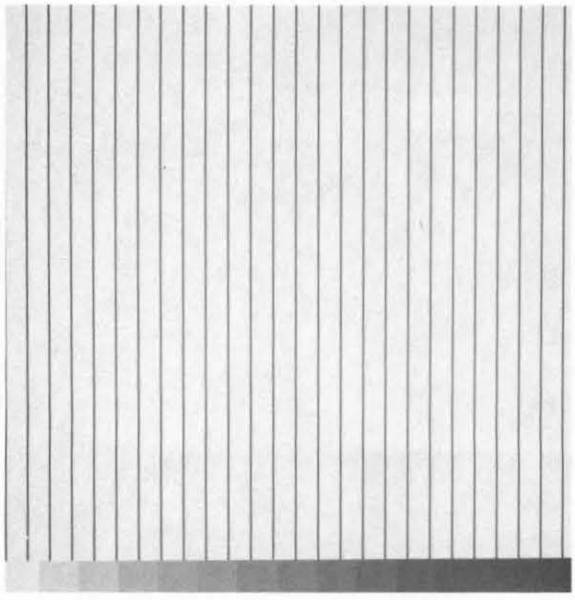


Figure 6.25 - Near 100% duty cycle rectangular wave grating.

gratings are varied in contrast horizontally and then surrounded by the masking stimuli. The reader should first cover up the masking stimulus and note the point of disappearance of each grating (waiting enough time to allow the adapted thresholds to settle!). Uncovering the masking stimuli shifts the threshold for that spatial frequency. The difference between the new and old thresholds gives the threshold elevation. Figures 6.26, 6.27 and 6.28 give images constructed with sine wave gratings consisting of the fundamental, second and third harmonics of the masking stimulus.

The reader should conclude from the above experiment that a high contrast, near 100% duty cycle rectangular wave grating strongly adapts and masks only the fundamental frequency simusoidal grating.

6.5 Brightness Perception Models

Lateral Inhibition. Perhaps the most successful model of neurophysical architecture to date is the lateral inhibition model for brightness perception. Mach (1868) pointed out how such an organization could be responsible for the response elicited by the visual system to the Mach band stimuli. Furthermore, the celebrated experiments of Hartline (Ratliff 1965) actually demonstrated electrophysiologically the existence of structures postulated by Mach over a century ago. Finally, the

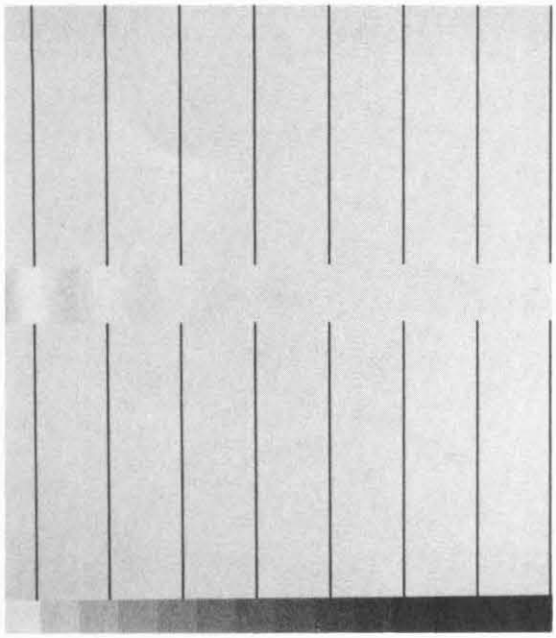


Figure 6.26 - Masking of fundamental by rectangular wave.

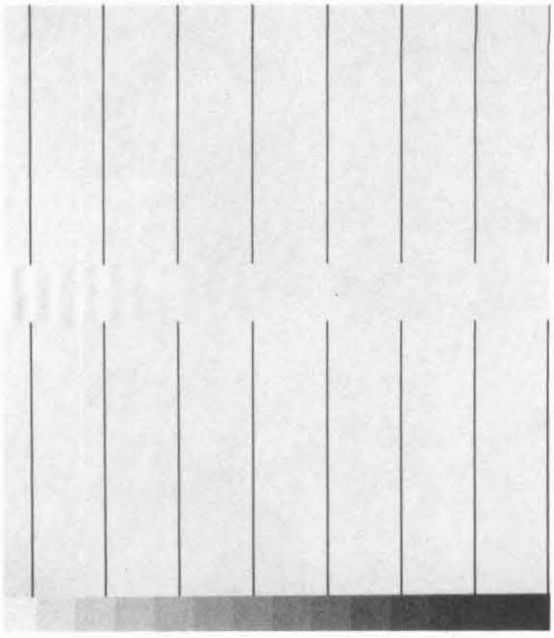


Figure 6.27 - Masking of third harmonic by rectangular wave.

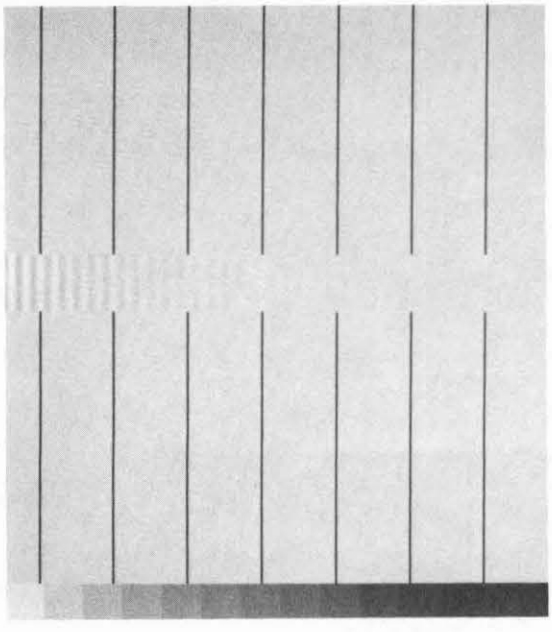


Figure 6.28 - Masking of fifth harmonic by rectangular wave.

usefulness of the model in predicting and explaining various psychophysical phenomena beyond the Mach band illusion has earned the lateral inhibition model a landmark position in the modelling of human vision.

A convenient realization of the lateral inhibition model is studied in Baudelaire (1973). This model was first given by Stockham (1968) and is shown in Figure 6.29. The filter modulation transfer function is given in Figure 6.30. A related detection model, called the single channel model, is shown in Figure 6.31. If the energy in the filtered signal exceeds the threshold for detectability, then the observer will indicate that he sees an image. This model allows one to measure the MTF by performing the single sinusoid detection experiment described above. Indeed, the MTF measured in this way is in good agreement with measurements of the MTF obtained in various other ways (see Baudelaire 1973).

We shall now evaluate the lateral inhibition model with respect to the phenomena discussed in the previous sections. Of course, with respect to the Mach band illusion, the lateral inhibition model elicits responses with qualitative and quantitative similarity to the actual brightness responses of the visual system. However, when the image possesses a luminance discontinuity, the lateral inhibition model predicts the strongest Mach band response at that edge. This is in strong disagreement with the

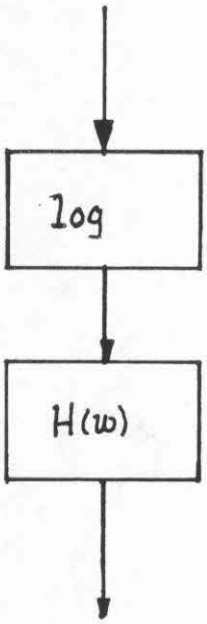


Figure 6.29 - Homomorphic lateral inhibition model.

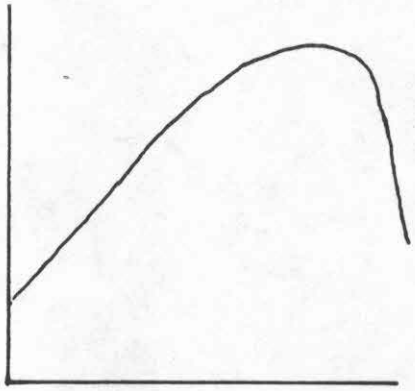


Figure 6.30 - Filter frequency response.

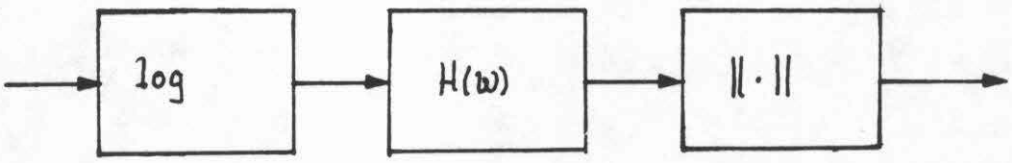


Figure 6.31 - Homomorphic lateral inhibition detection model.

brightness response of the visual system (where the Mach band disappears under these conditions). The edge scalloping illusion may be thought to possess an ostensible explanation by the lateral inhibition model but quantitative study shows that the model does not predict the actual appearance of the stepwedge. Furthermore, since the lateral inhibition model is local, that is, the brightness perception of a given point is unaffected by the luminance of points at distances of more than approximately 10-20 minutes of visual arc, the dependence of edge scalloping upon the number of edges visible remains unexplained. Again, the local nature of the lateral inhibition model prevents any explanation or prediction of the Cornsweet and Missing Fundamental illusions to appear plausible. As far as detection experiments go, the model agrees well with the single simusoid case. However, the principal reason that the single channel model does not enjoy a prominent status today is its complete failure to account for the results of the two sinusoid detection experiments as well as the masking and adaptation experiments described above (Graham and Nachmias 1971). Furthermore, as far as the new phenomena are concerned, the model does not correctly predict any responses. Note that the failure of the model, especially in predicting illusion softening of the Mach effect, is significant. We now describe alternative

models for brightness perception.

Retinex. The retinex model appeared first in Land (1964) and is closely related to a suggestion by Fry (1948). There have been a number of equivaluent formulations, the simplest of which have appeared in Horn (1973) and Baxter (1975). All these models attempt in various ways to throw away all contrast information not at an edge (C discontinuity). They were intended mainly for explaining the mechanism of the visual system that enables it to accurately judge surface reflectance or "lightness" independently of illuminance*. Recently Marr (1974) has proposed that there exist
physiological structures in the retina that implement the (Horn's) retinex model. Horn's and Baxter's retinex are shown in Figures 6.32 and 6.33. Both these models employ some ratio mechanism (implemented by differencing of the logarithm) followed by a threshold and a suitable reconstruction operator. Thus, when the luminance profile of an image varies slightly with position, the retinex will set that variation to zero. Only gross variations over small distances will exceed the threshold set in the middle stage of the retinex, thus affecting the ultimate output of the model, i.e. the proposed brightness

---

*The lateral inhibition model also does this. (See Stockham (1972).)

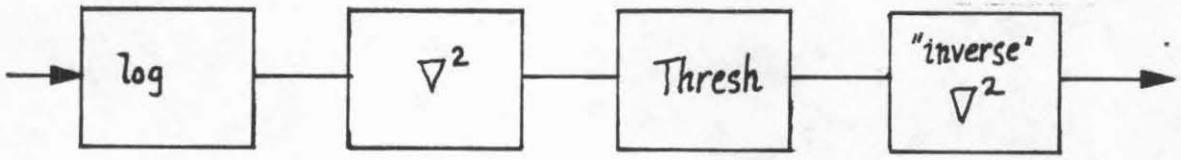


Figure 6.32 - Horn Retinex

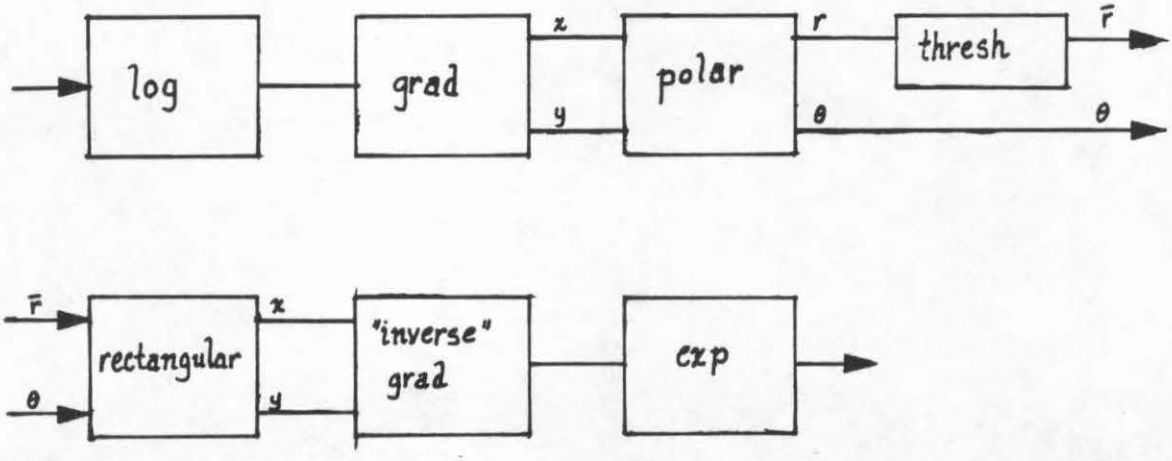


Figure 6.33 - Baxter Retinex.

response.

Let us now evaluate this model with respect to the phenomena discussed in previous sections. Among the first tests of the retinex in the perception of images was the response of the retinex to a certain subclass of images, the so-called Mondrians (Land and McCann 1971). These images are simply step function (piecewise constant) images multiplied by a slowly varying illuminance profile. The retinex does a reasonable job in recovering the step function image, however its performance outside of this artificial subclass is the key point that will interest us here. Considering the Mach band response and the edge scalloping response to a step wedge (a Mondrian!), the retinex model fails completely to match the response of the visual system. The Cornsweet and Missing Fundamental illusions are predicted successfully since the retinex computes the brightness of the various fields solely at the sharp contours between adjacent areas. That the Cornsweet illusion is predicted accounts for much of the appeal of this model. The retinex model is inconsistent with all the detection and adaptation experiment results, even in the single sinusoidal grating case. Only one possible MTF is given by the model which is dictated by the differencing operator. If the model is modified to compute a filtering operation similar to the lateral inhibiton model instead of a differencing operation, then

the retinex is converted to a model similar to the single channel detection model, thus not only performing adquately in the single sinusoid detection experiment, but also sharing the concomitant inadequacies of the single channel detection model. As far as the illusion softening phenomena go, the retinex fails completely. This is significant especially with the first example of the multiple (non-alternating) Cornsweet edges in Figure 6.14. Although the retinex correctly predicts the brightnes response to a single Cornsweet edge, it fails to predict the illusion softening. The new illusion presented in Section 6.4 also poses some difficulty for acceptance of the retinex as a valid model for the human visual system. Here, the Cornsweet-like brightness response fails to be produced by the retinex because the image contains no sharp discontinuities so that the retinex is unable to compute brightness differences.

Receptive Field/Spatial Filter. The receptive field and spatial frequency filter, or multiple channel detection model, are the current most popular models in the psychophysical community (Robson 1975, Campbell and Robson 1968). The receptive field model was motivated by the celebrated experiments of Hubel and Wiesel (1962), where the responses of individual neuronal units in the visual cortex of a cat were recorded for various visual stimuli.

It was found that the (simple) units corresponded to specific areas on the retina and that their spontaneous activity could be enhanced or inhibited by luminances within so-called excitatory and inhibitory regions. Further researchers found that the organization of these regions could be considered to be a main excitatory region surrounded by several lobes of alternating excitatory and inhibitory regions. These and other observations led Pantle and Sekuler (1968) and Campbell, et al. (1969) to postulate a spatial frequency filter interpretation for the receptive fields. The receptive field and spatial frequency filter are completely equivalent by the argument given in Section 5.3. These models may be schematized as in Figure 6.34. The logarithm may or may not be included in the detection model since for the detection stimulus we have

$$S_2: L_o(1+m\sin\omega x).$$

Taking the logarithm, we obtain

$$\log L_o(1+m\sin\omega x) = \log L_o + \log(1+m\sin\omega x)$$

$$\approx \log L_o + m\sin\omega x$$

where the last approximation holds for $m<<1$. This condition is almost always satisfied in a detection experiment. As Section 5.3 showed, the receptive field model can be considered to be simply a bank of filters. Thus the model is simply a bank of bandpass filters scaled according to the MTF of the visual system followed by a

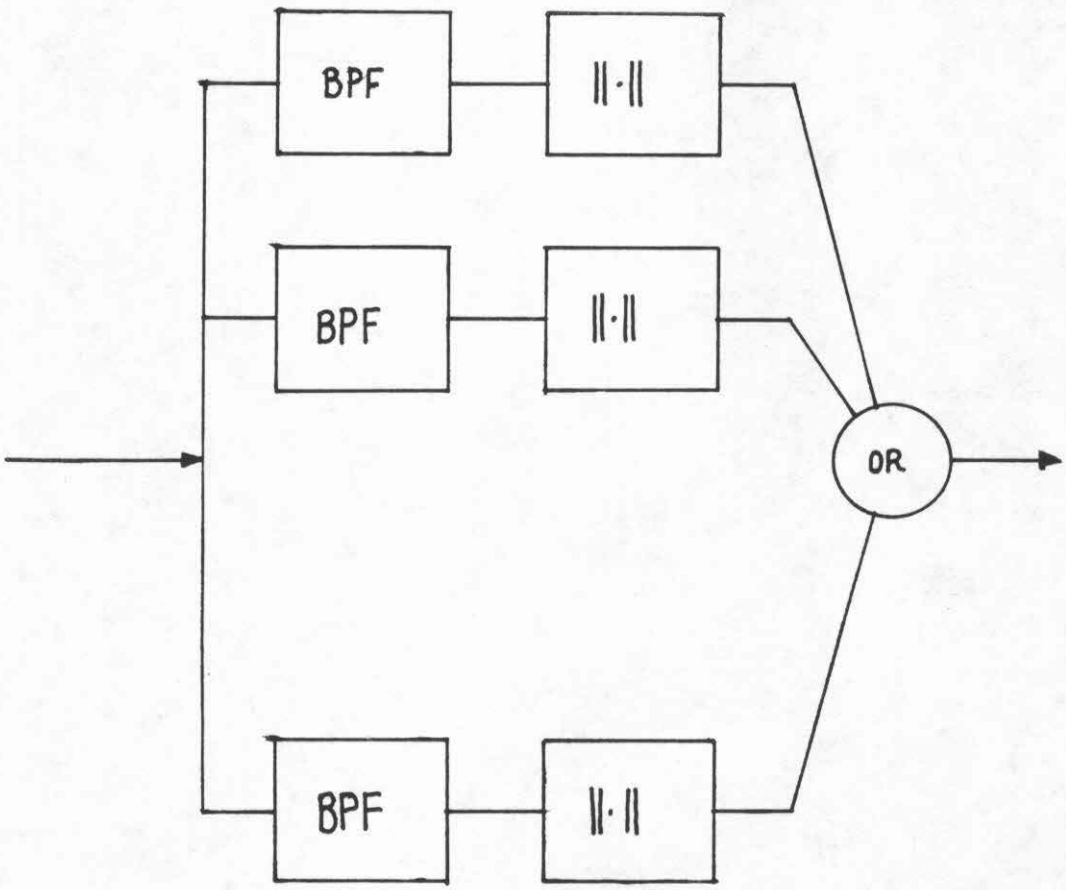


Figure 6.34 - Receptive field/spatial filter detection model.

norming operation. The real difference between the receptive field and spatial frequency detection models is in the norming operation. Campbell and Robson (1968) assume a norm that is L^2 in the spatial direction (x) and L^∞ in the frequency (ω). Mostafavi and Sakrison (1976) favor an L^p norm across space for $p \approx 6$ and L^∞ norm across frequency. Sachs, Nachmias and Robson use L^2 across space and $L^{7.8}$ across frequency (see Quick 1974, Graham 1977). In the receptive field model, the norm across space is $L^{p'}$ with p', a high number. In the above model structure, the free elements are the window function h (determining the spatial frequency filter profile k) and the norming operation. Controversies occur over the form that these two elements should take, but the basic model structure is generally agreed upon.

Let us now examine the performance of this model with regard to the above phenomena. Disregarding the norming operation and treating the outputs of the scaled frequency channels as the perceived image, the model collapses to the lateral inhibition model. It reverts automatically to the single channel model when presented with single sinusoid gratings. The Cornsweet and Missing Fundamental illusions are explained by a clever unconscious inference argument due to Campbell, et al. (1971). In this explanation, the Missing Fundamental is not noticed by the visual system since for low frequencies the fundamental of

a square wave is already below the threshold for detection. Thus the visual system cannot distinguish between the actual rectangular wave and a missing fundamental wave. By unconscious inference, when presented with such stimuli it chooses to perceive the square wave train. The argument is similar for the Cornsweet illusion. Sullivan and Georgeson (1977) have quantitatively tested this theory for conditions of varying temporal frequencies and mean luminance (under both these cases the low frequency thresholds shift considerably) and found excellent agreement between predictions of the theory and experimental fact. This explanation also accounts for the edge scalloping illusion: when only a few steps are uncovered, then the fundamental in the stepwedge is below its detection threshold so that the visual system chooses the perception of a step; when more edges are uncovered, the fundamental of the stepwedge is above threshold so that the high pass characteristic of the MTF can cause the scalloped perception (the visual system can no longer arbitrarily choose the step brightness response because it no longer produces the same signal as the stepwedge whose fundamental is below detection threshold). All of the detection experiment results are accounted for in this model, but since this is primarily a detection model, no adaptation phenomena are correctly predicted. The next

model discussed incorporates adaptaion into this model structure. Finally, illusion softening of the Cornsweet response is readily explained in an argument quite similar to the edge scalloping discussion. Softening of the Mach response fails to be predicted by this model.

Spatial Filter/AGC. An attempt to incorporate the adaptation response into the spatial frequency filter model structure is given in Baxter (1975), see Figure 6.35. This model has been implicit within all spatial frequency filter models from the very beginning (Blakemore and Campbell 1969). Because this model structure is an extension of the previous structure, the successes of the previous model are enjoyed by the present model. Furthermore, the adaptation phenomena discussed in previous sections can now be considered. First, Blakemore and Campbell's (1969) single sinusoid adaptation experiment results are closely accounted for by the present model. When adapting to a two sinusoid grating or a square wave grating (Tolhurst 1972), however, the predictions suffer in accuracy. Recall that under these conditions the third harmonic is not nearly as effective an adapting stimulus as when presented in isolation. The present model fails to account for this effect in any way. Tolhurst took the above effect as sufficient reason to consider inhibition between the channels (during

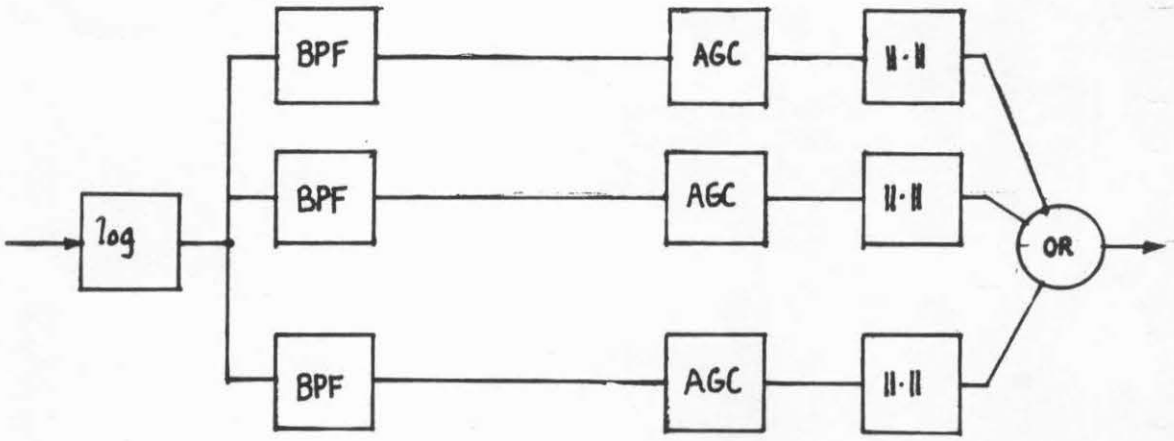


Figure 6.35 - Spatial filter/AGC adaptation model.

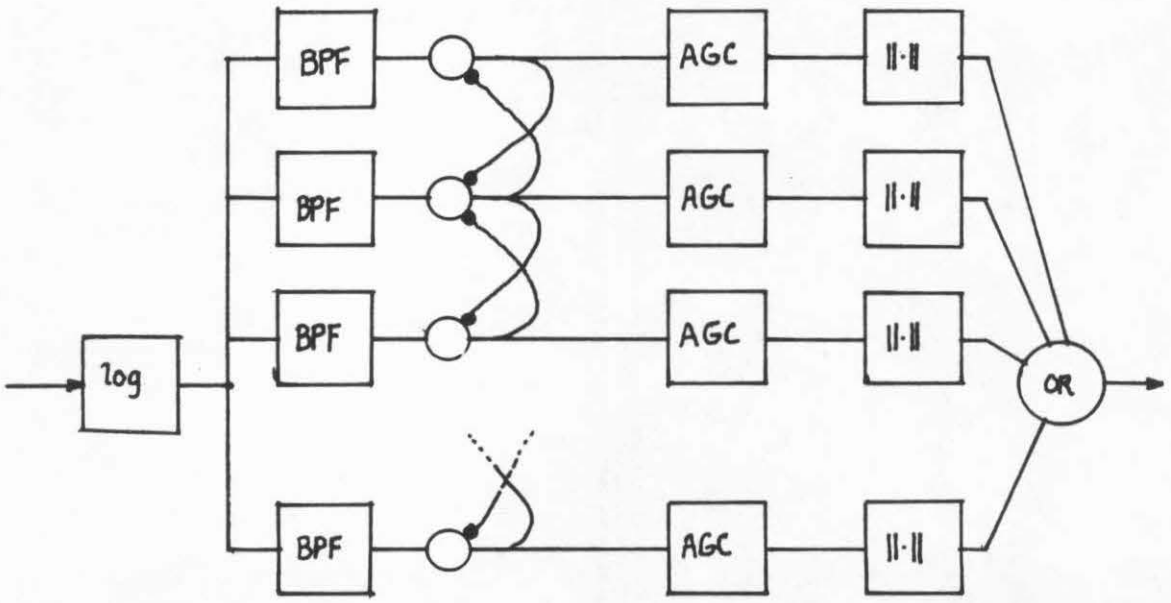


Figure 6.36 - Tolhurst adaptation model.

adaptation only). Thus, Tolhurst's model appears as in Figure 6.36. The reader at this time should recall the discussion on ad hoc modelling in Chapter 1. This steady progression from a single channel, to multiple channel, to multiple channel with AGC, to finally multiple channel with AGC and suprathreshold inhibition between channels model is precisely the situation discussed there. Finally, the results from Maudarbocus and Ruddock must be considered. The dependence of adaptation upon mean luminance, they argue, is evidence for a nonlinearity prior to the spatial frequency filtering stage. Note that the Tolhurst model does not possess this very important (receptor) nonlinearity, consequently many of his calculations may be under suspicion. The next chapter considers this matter more thoroughly. The second result of Maudarbocus and Ruddock, that an extremely high contrast sinusoid does not cause significant changes in the shape of the threshold shift curves, places another doubt as to the validity of the spatial frequency filter adaptation model (with receptor nonlinearity, with or without inhibition between channels). As they point out, high contrast adapting gratings, when passed through the receptor nonlinearity, no longer retain their pure character; that is, higher harmonics are present in the processed signal not present in the original. Thus, given the spatial frequency filter model, significant threshold

shifts at the second harmonic of a high contrast adapting grating should be expected. None occur. Consideration of the remaining new phenomena also presents difficulties for this model. The illusion softening effects for the Cornsweet illusion have been previously explained. For the Mach band response, adaptation of the higher frequencies because of the surrounding edges will lessen the high pass characteristic of the MTF, thus reducing the Mach band. This qualitative explanation requires further quantitative study. Adaptation to the near 100% duty cycle rectangular wave train poses the principal challenge to the present model. As Fourier theory predicts, the energy in the fundamental decreases with the deviation of the duty cycle from 50%. Note that a rectangular wave retains its shape through the receptor nonlinearity; thus it does not matter whether the nonlinearity is included or not. If there is no inhibition between channels, the model predicts that only a weak adaptation to the fundamental equal to the adaptation at all harmonics will result. With inhibition between channels, the adaptation response will be predicted to be even smaller. These predictions are directly contrary to the facts.

For these reasons, we believe that although the multichannel detection model is valid, any prediction made by its adaptation/AGC extension should be viewed with critical suspicion.

CHAPTER 7

THE NEW MODEL

7.1 Introduction

In the last chapter we found that many ostensible models for the visual system suffer from some defect, viz. an inability to produce responses similar to those of the visual system when presented with certain stimuli. The most successful model to date is the multichannel detection model, which accounts for all the visual phenomena except the adaptation response. Extensions of this multichannel model to include adaptation have not been nearly so successful, as we have seen in Chapter 6. This chapter will present a new extension of the multichannel detection model that produces adaptation responses in agreement with the visual system to all the stimuli described in the previous chapter.

In the next section we will carefully analyze the initial stages of the detection model in order to better understand the mechanism responsible for the failure of present adaptation models to accurately predict visual system behavior. Specifically, the effect of the initial receptor nonlinearity upon the adapting stimuli is

calculated. Next, in Section 7.3 we discuss a particular estimation scheme popular in nonlinear modulation theory. We will propose this scheme as a basis for a model of the adaptation response of the visual system. In Section 7.4, we discuss the performance of these models tested against the stimuli in Chapter 6. Finally, in Section 7.5 the response of the full two-dimensional model to real world images will be presented and the results of a texture experiment will be described.

7.2 An Analysis of the Adaptation Experiment

Under a detection experiment we showed in Section 6.5 that, if the receptor nonlinearity were a logarithm, then it could be safely ignored. This calculation was an example of how small perturbations may effectively linearize nonlinear phenomena. Davidson (1968) extends these kinds of arguments to other nonlinearities. Thus a theoretical setting for a linear multichannel detection model can pin down the structure quite precisely, i.e. the Mandala transform and perturbational argument give the multichannel filter bank directly. For the adaptation experiment, however, this situation changes entirely. Because the adapting stimuli frequently have large contrasts, the nonlinearity may no longer be ignored. Tolhurst (1972) does not take this observation into account when performing his calculations.

Evidence presented in Burton (1973) shows that the nonlinearity observed in the visual system occurs before the mechanisms that define the adaptation response. Specifically, two sinusoidal gratings with spatial frequencies beyond the resolution of the visual system are simultaneously projected on the retina. These two sinusoids differ slightly in spatial frequency such that the difference of the two frequencies lies within the range to which the visual system is responsive. If the nonlinearity in the visual system occurs after the stage that determines the adaptation response, then no rise in threshold should occur for any spatial frequency (remember that the two sinusoids have frequencies beyond the limit of resolution of the visual system). This is not the case, however. A threshold rise at the difference frequency indicates that some nonlinearity is responsible for the mixing of the two sinusoids to produce a beat frequency. Thus there is a nonlinearity in the visual system before the adaptation response mechansm. This nonlinearity is usually associated with the receptor neuron charged with the responsibility for transduction of light energy to neural impulses. There have been several proposals for the form that this nonlinearity should take. We shall proceed with the analysis with two forms in mind. The first, of course, is the logarithm (Baudelaire 1973). The second is due to Alpern, Rushton and Torii (1970). It

is $\frac{x}{x+I}$, where I is some experimentally determined constant (about $10^{3.27}$ troland).

We shall investigate the Fourier series development of adapting stimuli passed through two proposed receptor stimuli. Let the adapting stimulus be

$$L_0(1+m\sin x).$$

We wish to calculate the quantities

$$a_n = \int_0^{2\pi} \log(L_0(1+m\sin x))\, e^{-inx}\, dx$$

$$b_n = \int_0^{2\pi} \frac{L_0(1+m\sin x)}{L_0(1+m\sin x)+I}\, e^{-inx}\, dx.$$

We first proceed with the calculation of a_n. Since we are interested in a_n for n=1,2,... the reader can easily see that L_0 does not affect any non-DC term and may be neglected. Making the substitution $\zeta = e^{ix}$ and expressing the term as a line integral on the unit circle,

$$\begin{aligned} a_n &= \int_0^{2\pi} \log(1+m\sin x)\, e^{-inx}\, dx \\ &= \oint \log\left(1+\frac{m}{2i}\left(\zeta-\frac{1}{\zeta}\right)\right) \frac{1}{i\,\zeta^{n+1}}\, d\zeta \\ &= \oint \log\left(\frac{\zeta^2-\frac{2i}{m}\zeta-1}{\frac{2i}{m}\zeta}\right) \frac{1}{i\,\zeta^{n+1}}\, d\zeta \\ &= \oint \left\{\log(\zeta-\gamma_1)+\log(\zeta-\gamma_2)-\log\left(\frac{2i}{m}\zeta\right)\right\} \frac{1}{i\,\zeta^{n+1}}\, d\zeta. \end{aligned}$$

This last equality follows from the factorization of the quadratic and using the elementary properties of the logarithm.

We have set γ_1, γ_2 equal to the roots of the equation.

$$\gamma_1 = -\frac{i}{m}+\sqrt{1-\frac{1}{m^2}}$$

$$\gamma_2 = -\frac{i}{m}-\sqrt{1-\frac{1}{m^2}}$$

Of course, the contrast m is always restricted to $m<1$, thus γ_1, γ_2 are evidently pure imaginary quantities.

We must now evaluate the three integral quantities. Set

$$g_1(\zeta) \equiv \log(\zeta - \gamma_1)$$

$$g_2(\zeta) \equiv \log(\zeta - \gamma_2)$$

$$g_3(\zeta) \equiv \log(\zeta)$$

$$g_4(\zeta) \equiv \log\left(\frac{2i}{m}\right) = z_0 .$$

Then g_1, g_2, g_3 are multiple valued functions with branch points $\gamma_1, \gamma_2, 0$ and cuts shown in Figures 7.1a,b,c.

Thus the line integrals are given by

$$I_1 = 2\pi\{1 + g_1^{(n)}(0)\}$$

$$I_2 = 2\pi g_2^{(n)}(0)$$

$$I_3 = \frac{2}{\pi}\frac{(-1)^n}{n}$$

$$I_4 = 0 .$$

The first two integrals are calculated with the aid of the Cauchy integral formula along the contours shown in Figure 7.1. The third and fourth integrals are calculated directly.

Thus we have that

$$a = I_1 + I_2 + I_3 + I_4$$

$$= 2\pi - 2\pi\frac{1}{\gamma_1^n (n-1)!} - 2\pi\frac{1}{\gamma_2^n (n-1)!} - \frac{2}{\pi}\frac{(-1)^n}{n} .$$

But because $\gamma_1^n \gamma_2^n = (-1)^n$, we have

$$a_n = 2\pi - \frac{2\pi}{(n-1)!}\left[\frac{\gamma_1^{2n} + (-1)^n}{\gamma_1^n}\right] - \frac{2}{\pi}\frac{(-1)^n}{n} .$$

The dependence of a_n upon m is shown in Figure 7.2 for $n=1,2,3$.

To calculate b_n, we take a slightly different

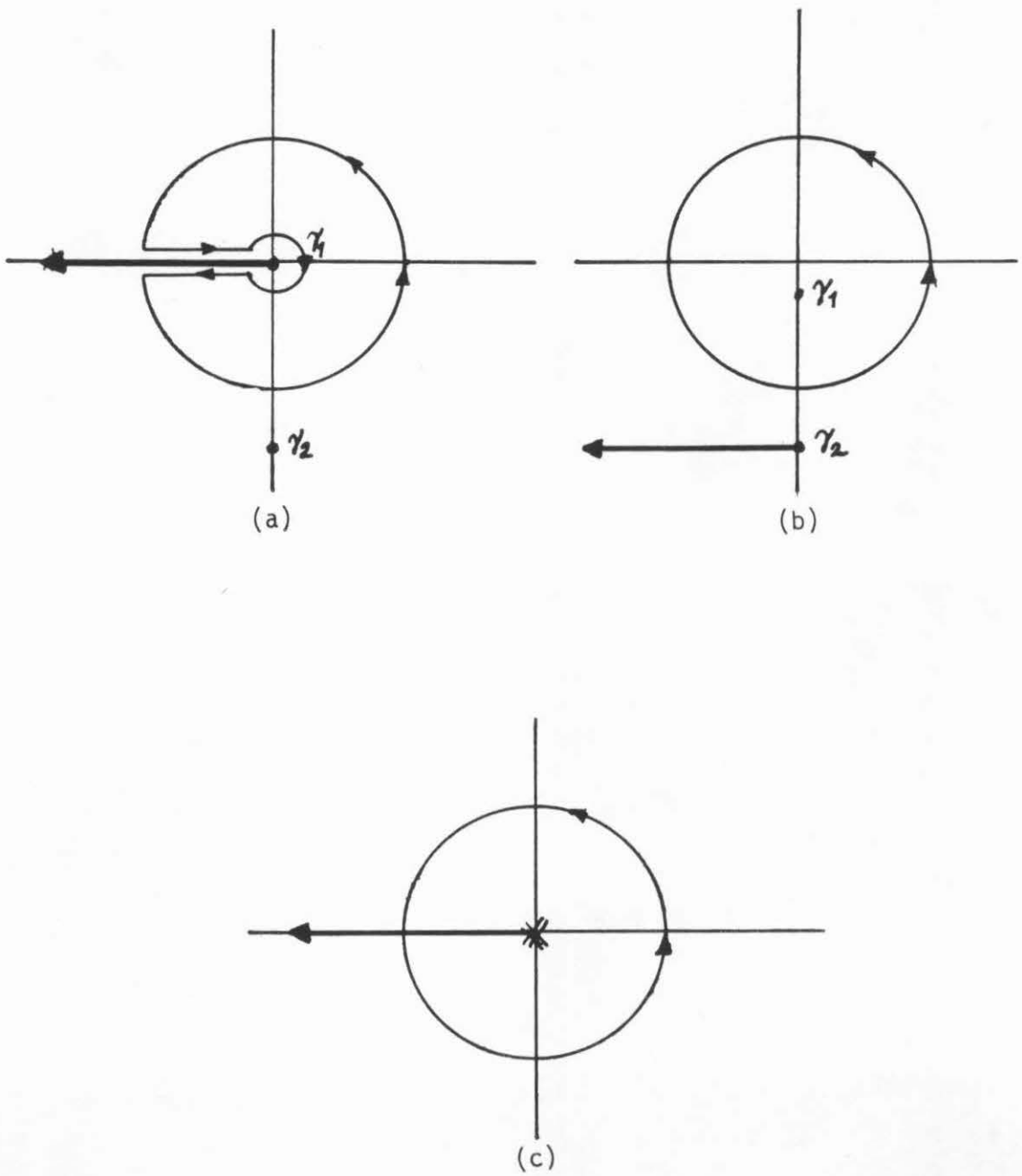


Figure 7.1 - Integration contours.

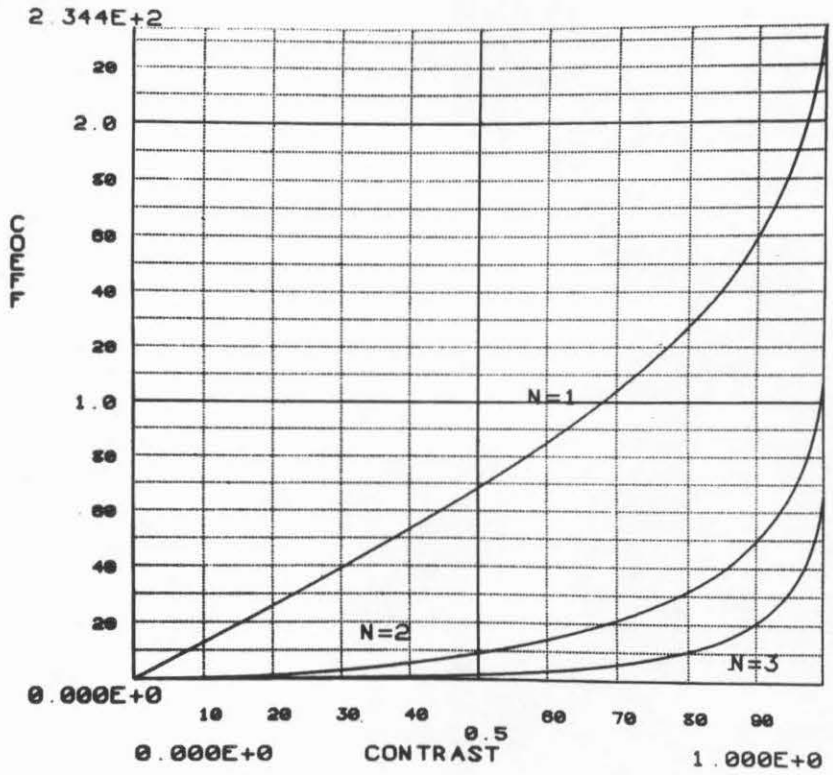


Figure 7.2 - a_n vs. contrast.

approach. First rewrite the adapting stimulus

$$(7.2.1)\qquad \frac{L_0(1+m\sin\chi)}{L_0(1+m\sin\chi)+I} = \frac{1+m\sin\chi}{m\sin\chi+\alpha} \qquad m<1,\ \alpha>1$$

where $\alpha=\frac{I}{L_0}+1$. We now rewrite $\sin x=\frac{1}{2i}(\zeta-\frac{1}{\zeta})$ for $\zeta=e^{ix}$.

$$\frac{1+\frac{m}{2i}\left(\zeta+\frac{1}{\zeta}\right)}{\frac{m}{2i}\left(\zeta+\frac{1}{\zeta}\right)+\alpha} = \frac{\zeta^2+\frac{2i}{m}\zeta-1}{\zeta^2+\frac{2i\alpha}{m}\zeta-1} = \frac{A}{\zeta-\gamma_3}+\frac{B}{\zeta-\gamma_4}$$

The last equality is the partial fraction expansion with residues

$$A=\frac{(\gamma_3-\gamma_4)(\gamma_3-\gamma_2)}{(\gamma_3-\gamma_2)}$$

$$B=\frac{(\gamma_4-\gamma_1)(\gamma_4-\gamma_2)}{(\gamma_4-\gamma_3)}$$

where γ_i are the zeroes and poles of the adapting stimulus (7.2.1),

$$\gamma_1=-\frac{i}{m}+\sqrt{1-\frac{1}{m^2}}$$

$$\gamma_2=-\frac{i}{m}-\sqrt{1-\frac{1}{m^2}}$$

$$\gamma_3=-\frac{i\alpha}{m}+\sqrt{1-\frac{\alpha}{m^2}}$$

$$\gamma_4=-\frac{i\alpha}{m}-\sqrt{1-\frac{\alpha}{m^2}}.$$

We now wish to expand the partial fraction expression into a Laurent series convergent in an annulus containing the unit circle. Then substituting $=e^{ix}$ will give the Fourier series development of the adapting stimulus. From the standard constraints on the parameters in (7.2.1), it is evident that $|\gamma_3|$ and $|\gamma_4|$ form the outside and inside radii of such an annulus (remember that $\gamma_3\gamma_4=-1$). We can readily expand the partial fraction expression by a geometric series to yield

$$-\frac{1}{\gamma_3}\frac{(\gamma_3-\gamma_1)(\gamma_3-\gamma_2)}{(\gamma_3-\gamma_4)}\sum_{n=0}^{\infty}\left(\frac{\zeta}{\gamma_3}\right)^n+\frac{(\gamma_4-\gamma_1)(\gamma_4-\gamma_2)}{(\gamma_4-\gamma_3)}\frac{1}{\zeta}\sum_{n=0}^{\infty}\left(\frac{\gamma_4}{\zeta}\right)^n$$

Thus the Fourier coefficients are

$$b_n = \begin{cases} \dfrac{-(\gamma_3\gamma_1+1)(\gamma_3-\gamma_1)}{} \dfrac{1}{\gamma_3^n} & n=0,1,2,\ldots \\ (-1)^{n+1} \dfrac{(\gamma_3\gamma_1+1)(\gamma_3-\gamma_1)}{\gamma_1\gamma_3(\gamma_3^2+1)} \dfrac{1}{\gamma_3^{n+1}} & n=-1,-2,-3,\ldots \end{cases}$$

The dependence of b_n upon m is shown in Figure 7.3.

Now let us consider the spectrum of a square wave passed through these nonlinearities. It is evident that the result of passing a square wave through a memoryless nonlinearity is simply another square wave. If an adapting stimulus is of the form

(7.2.2) $$L_0(1+m_a s(x))$$

where

$$s(x) = \begin{cases} 1/2 & n \leq x < n+ \ , \ n \in Z \\ -1/2 & n+\frac{1}{2} < x < n+1, \ n \in Z \end{cases}$$

then its mean luminance is L_0 and its contrast is m. Now the Fourier series of such a stimulus is

$$\sum_{n=-\infty}^{\infty} c_n e^{inx} = \sum_{n=-\infty}^{\infty}{}' \frac{2m L_0}{i\pi n} e^{inx} + L_0 .$$

Passing this stimulus through the logarithm gives a square wave:

$$L_0'(1+m_a' s(x))$$

where $L_0' = \log(L_0\sqrt{(1-m_a)})$, $m_a' = \dfrac{\log\left(\frac{1+m}{1-m}\right)}{\log(L_0^2(1-m^2))}$. Thus the Fourier coefficients for a square wave passed through a nonlinearity are simply

$$d_n = \begin{cases} \frac{2}{i\pi n} L_0' m_a' & n \neq 0 \\ L_a' & n=0. \end{cases}$$

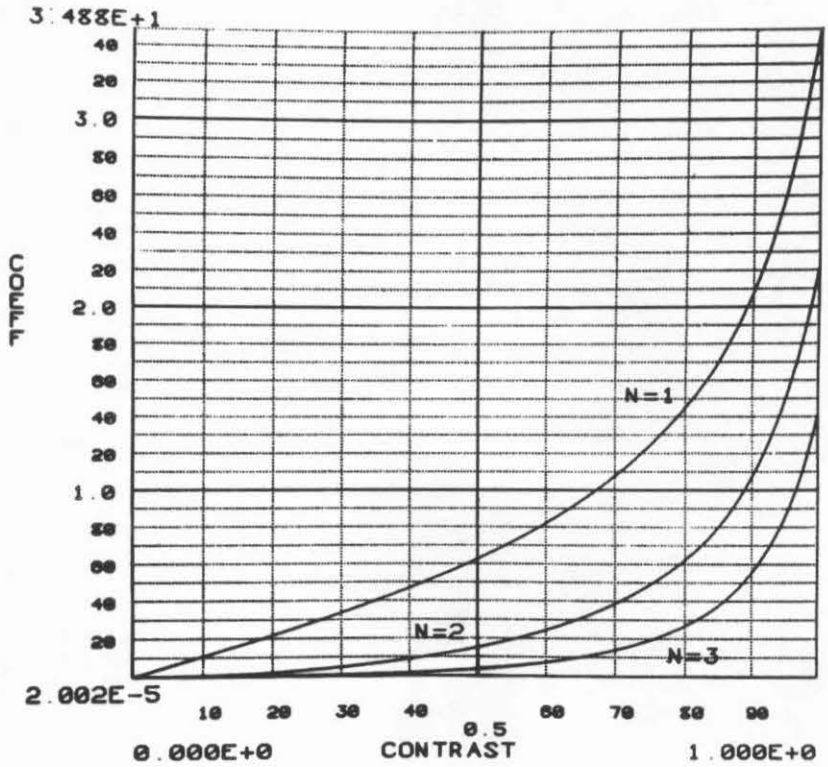


Figure 7.3 - b_n vs. contrast.

The variation of d_n versus m is shown in Figure 7.4 for $L_0=1$. If we pass (7.2.2) through the nonlinearity of Alpern, Rushton and Torii (1970), we get:

$$L_0'(1+m_a's(x))$$

where $L_0'=\frac{\alpha-m_a^2}{\alpha^2-m_a^2}$, $m_a'=\frac{(\alpha-1)m_a}{\alpha-m_a^2}$. The behavior of the coefficients is shown in Figure 7.5a,b for $\alpha=1.1$ and $\alpha=10$, viz. for high and low mean luminances respectively.

Let us now discuss how the spatial filter/AGC adaptation models behave with respect to the stimuli passed through the nonlinearities discussed in this chapter. First of all, one should note that the shape of the a_1, b_1 curves imply that the adaptation response at the fundamental would be a non-monotonic function of the adapting contrast and that the behavior of these curves indicate that the choice between these nonlinearities is not critical for our concerns. This is in disagreement with the facts established in Section 6.3. Furthermore, as the a_2, a_3, b_2, b_3 coefficients indicate, significant second and third harmonic energies are present in the signal passed through the nonlinearity. A proponent of these models should predict that, at high contrasts, significant adaptation to higher harmonics would occur. Again, these effects are inconsistent with observations of the visual system. Finally, as a perusal of the d_n, e_n coefficients would reveal, Tolhurst's calculations must be rejected. This is because the nonlinearities cause much

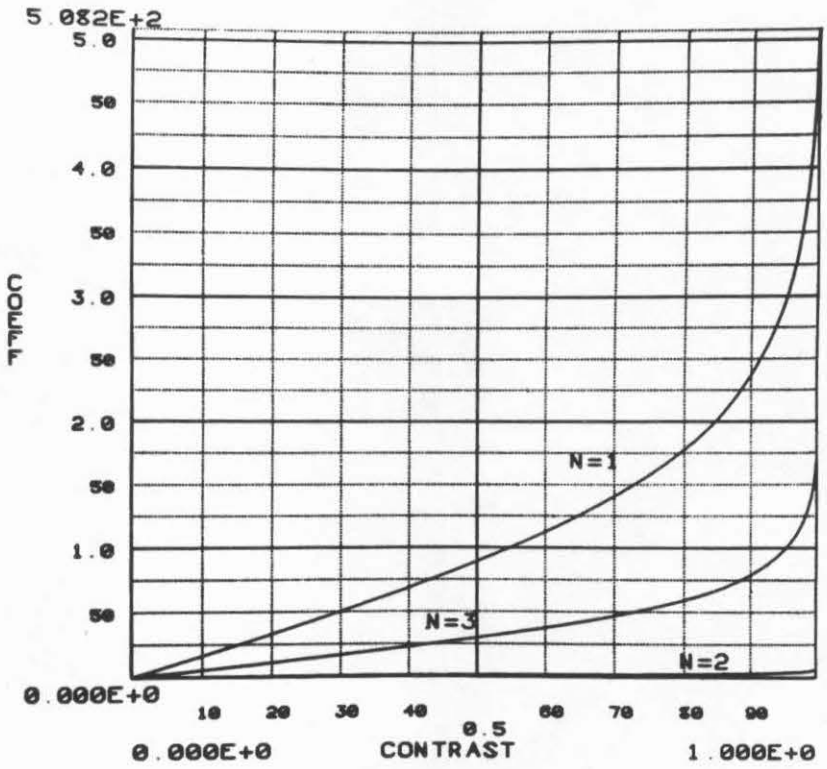
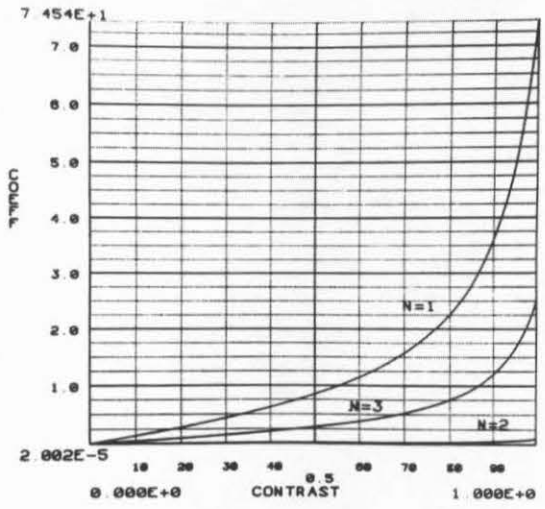
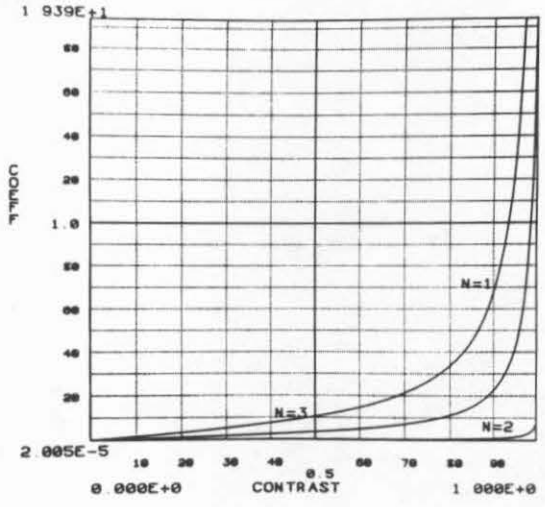


Figure 7.4 - d_n vs. contrast.



(a)



(b)

Figure 7.5 - e_n vs. contrast: (a) Mean luminance = 1, (b) Mean luminance = 10.

higher contrasts in the square wave than in Tolhurst's analysis. His effect, however, remains in force -- indeed, an even higher attenuation of the expected adaptation response would be required from his proposed mutual inhibition scheme. Because of the similarity in these curves between these two nonlinearities, we shall restrict our attentions to the first, the logarithm.

7.3 A Maximum Likelihood Estimator

From the above discussion, it is clear that no linear processor acting upon the receptor nonlinearity output signal will be able to model the adaptation mechanism of the human visual system. We know that this nonlinearity does occur as one of the first components of the visual system signal processing chain, or at least occurs prior to the adaptation stage (Burton 1973). Thus we need to search for nonlinear processors.

It would be most desirable in terms of the symmetry theory to be able to deduce a structure for the adaptation stage through analysis of a sufficiently rich symmetry group. Unfortunately we have not been able to find such a group at the time of the present writing. Our inability to find such a group is due, in part, to the fact that adaptation is a quasi-temporal offset. That is, the adaptation experiment is not performed under stationary conditions but relies upon a change in the stimulus

presented to the observer. Thus, a group able to shed light upon the adaptation mechanism may be obtained only when temporal considerations enter into the overall analysis, an area that we have at the outset excluded from our investigation. We hope that at some future time an adaptation group may be found which will allow a formal analysis of the adaptation mechanism.

In the development below, we need to shift from formal modelling to ad hoc modelling, in order to proceed without knowledge of the appropriate symmetry group.

Intuitively, one may invent any number of schemes to estimate the contrast of a sinusoidal grating that has been passed through a known memoryless nonlinearity h. For example, if the frequency ω, phase (assume 0) and mean luminance L_0 of the sinusoidal grating are known, then the contrast m can be estimated from the received signal

$$r(x)=h(L_0(1+m\sin\omega x))$$

simply by choosing a point x_0 and calculating

$$\hat{m}=\left\{\frac{h^{-1}(r(x_0))}{L_0}-1\right\}\frac{1}{\sin\omega x_0}.$$

Furthermore if L_0 and ϕ are unknown then for a given ω one can, by choosing two (or more) points, construct a number of formulas similar to the above.

There are a number of shortcomings inherent within these relatively naive first attempts. First, we would like to use the entire received signal to calculate the estimate $\hat{m}$, as opposed to the above methods which only use

either a single point or a few points of the received signal. Second, if the received signal is not known exactly, but only approximately known with some uncertainty (as is the case in most physical situations due to measurement error or response noise), then the estimate $\hat{m}$ may become very unstable. Both of these difficulties may be overcome by the use of statistical methods. Because these statistical methods quite naturally process the entire signal in calculating an estimate $\hat{m}$, the first difficulty is overcome. The second is minimized by choosing a "best" (in some suitable sense) estimate $\hat{m}$ given the data received.

We will restrict our discussion to a particular technique, the so-called maximum likelihood estimation method (ML or MLE method). After the development of this method we shall mention the relationship between the results obtained for the contrast estimation task by the maximum likelihood method and those of other statistical methods.

The derivation sketched here can be found in van Trees (1968, p. 273). In the MLE method, the received signal is described by a joint probability density derived from the model depicted in Figure 7.6. For a given contrast m and phase ϕ the sinusoidal grating $L_o(1+m\sin\omega x+\phi)$ is generated. This signal is passed through a memoryless nonlinear transformation, which is

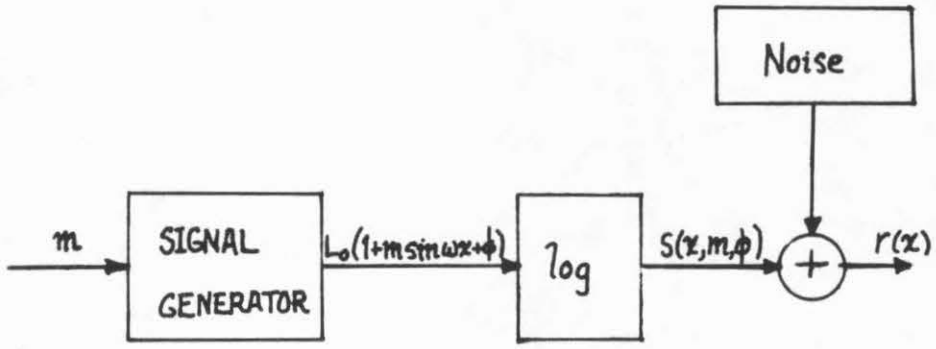


Figure 7.6 - Signal production model for MLE.

log(·). The maximum likelihood estimator observes a version r(x), corrupted by additive Gaussian noise with independent increments, of the nonlinearity output $s(x;m,\phi)$. The estimator must then calculate an estimate $\hat{m},\hat{\phi}$ of the original contrast m and phase ϕ. Thus the description of the received signal is a joint probability density given by

$$(7.3.1) \qquad p(R_1,\ldots,R_n|m,\phi)=\prod_{i=1}^{n}\frac{1}{\sqrt{\pi N_0}}\exp\left(-\frac{[R_i-s(x_i;m,\phi)]^2}{N_0}\right)$$

where N_0 is the standard deviation of the additive noise, $s(x_i;m,\phi)$ are samples of the signal output of the nonlinearity at the arbitrary points x_i, and R_i is the value of the received signal at the point x_i. Viewed as a function of $R_1,\ldots,R_n$ with fixed m,ϕ, this function p expresses (for a given m,ϕ) the probability of receiving a signal with values $R_1,\ldots,R_n$ at the sample points x_i. We may, however, give this function a different interpretation that will be very useful. Under our scheme the received signal r(x) is known but the actual values of m,ϕ (say m_o,ϕ_o) are unknown, hence the actual probability density $p(R_1,\ldots,R_n|\, m=m_o,\phi=\phi_o)$ is unknown. (Although of course the conditional density, that is, the functional dependence of p on m,ϕ is known.) For a given received signal r(x), let the values at sample points $x_1,\ldots,x_n$ be given by $\bar{R}_1,\ldots,\bar{R}_n$. Then p viewed as a function of m,ϕ given values $\bar{R}_1,\ldots,\bar{R}_n$ is called the <u>likelihood functional</u> for m and ϕ, viz.

(7.3.2) $$\Lambda[m,\phi;\bar{R}_1,\ldots,\bar{R}_n]=p(\bar{R}_1,\ldots,\bar{R}_n|\,m,\phi).$$

This function expresses the likelihood that a given value of m,ϕ generates the samples $\bar{R}_1,\ldots,\bar{R}_n$ under the situation depicted by Figure 7.6. The maximum likelihood method chooses as estimates the values $\hat{m},\hat{\phi}$ that maximize the likelihood Λ, i.e. the values most likely to produce the received signal distribution seen.

In order to obtain an equation for the ML estimate we first note that the abscissa of the maximum of a given function is unmoved if that function is scaled by a positive constant. That is, if f(x) is maximized at x^* then cf(x) has a maximum at x^* also (for $c>0$). Furthermore if g is a strictly monotonically increasing function then g(f(x)) possesses a maximum at x^* if and only if f(x) does. Thus making a monotonic transformation of the function does not alter the place at which the maximum occurs. Now, substituting Equation 7.3.1 into 7.3.2 gives us

$$\Lambda(m,\phi;\bar{R}_1,\ldots,\bar{R}_n)=\frac{1}{(\pi N_0)^{\frac{n}{2}}}\exp\Big(-\sum_{i=1}^{n}\frac{[\bar{R}_i-S(x_i;m,\phi)]^2}{N_0}\Big).$$

In light of the above discussion, we may ignore the leading constants and take logarithms to obtain a functional whose maximum coincides with that of Λ.

$$\Lambda^*=N\log[(\pi N_0)^{\frac{n}{2}}\Lambda]=-\sum_{i=1}^{n}[\bar{R}_i-s(x_i;m,\phi)]^2$$

where Λ^* is often called the log likelihood functional. For the limiting case of a continuum of points we refer the reader to van Trees (1968), where the development

results in

$$\Lambda^* = -\int_{-\infty}^{\infty} [r(x) - s(x;m,\phi)]^2 \, dx.$$

To obtain a necessary condition for the maximum we take the gradient of Λ (considered as a function of m and ϕ) and set it to zero to obtain the maximum likelihood equation.

(7.3.3) $$\nabla \Lambda^* = \int_{-\infty}^{\infty} 2\{r(x) - s(x;m,\phi)\} \nabla s(x;m,\phi) \, dx = 0.$$

Note that this equation is a system (∇s is a vector quantity) of (possibly) nonlinear equations.

Let us now apply these equations to the contrast estimation problem at hand. The log likelihood functional becomes

$$\Lambda^* = -\int_{-\infty}^{\infty} [r(x) - \log L_0(1 + m\sin\omega x + \phi)] \, dx$$

and the ML equation 7.3.3 becomes

(7.3.4a) $$\int_{-\infty}^{\infty} \{r(x) - \log L_0(1 + m\sin\omega x + \phi)\} \frac{\sin(\omega x + \phi)}{1 + m\sin(\omega x + \phi)} \, dx = 0$$

(7.3.4b) $$\int_{-\infty}^{\infty} \{r(x) - \log L_0(1 + m\sin\omega x + \phi)\} \frac{m\cos(\omega x + \phi)}{1 + m\sin(\omega x + \phi)} \, dx = 0.$$

A typical log likelihood function appears in Figure 7.7 where $L_0 = 1$, $\phi = 0$. The abscissa varies with $\hat{m}$. The actual contrast of the grating is $m_0 = .5$. Note that the maximum likelihood occurs at $\hat{m} = .5$. In Figure 7.8 we have solved the ML equation for a series of gratings whose contrasts are between $0 \leq m < 1$. The abscissa indexes the contrasts of the gratings, the ordinate m. Note that $\hat{m} = m$, i.e. the estimator (in the absence of noise) is completely accurate.

The relationships of the MLE technique to other

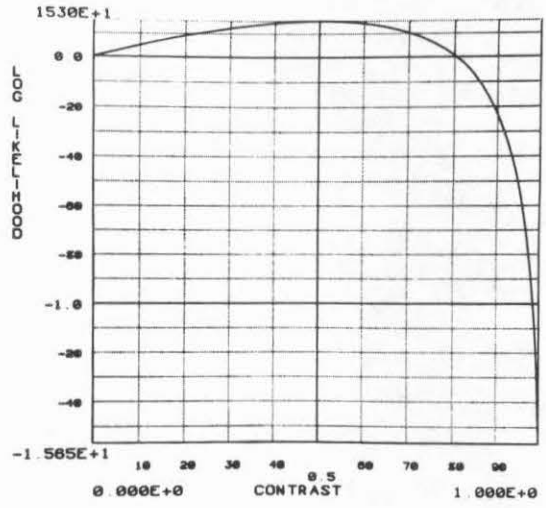


Figure 7.7 - Log likelihood functional.

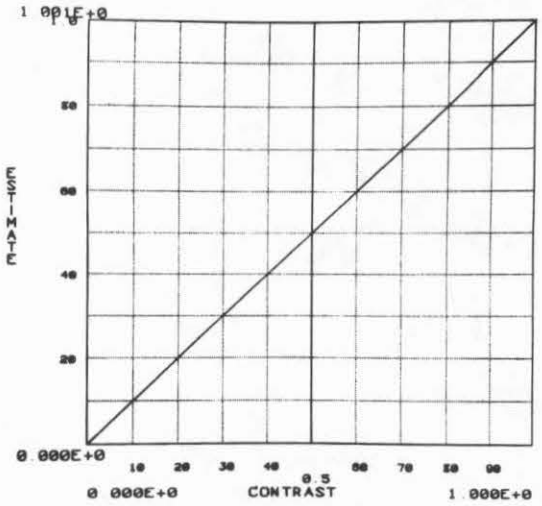


Figure 7.8 - MLE contrast estimate vs. actual contrast.

statistical techniques in this application are quite close. For example, those familiar with nonlinear least squares will recognize the ML Equations 7.3.4 as the solution minimizing the mean square error. For a Bayes estimation scheme it is easy to see that Equation 7.3.4 is also the optimum Bayes estimate given that the a priori probability density is not strongly concentrated about any particular value, i.e. that one is not predisposed toward any chosen contrast value. Furthermore, given rather reasonable restrictions on the risk function (see van Trees), the road always leads to Equations 7.3.4. Notice further that under a linear modulation scheme Equation 7.3.3 is simply a correlation operation, thus the MLE method could be said to be one generalization of correlation to nonlinear situations. Furthermore, for low contrasts m, the ML equations 7.3.4 collapse to a correlation operation, that is, correlation is a first order approximation to the operation performed in 7.3.4. It is these observations that lead us to believe that the Equations 7.3.4 represent a more fundamental result than their rather arbitrary derivation would at first sight indicate.

How our development parallels the well-known theory of ideal observers so popular in the current psychophysical literature should be obvious to those familiar with its details, cf. Green and Swets (1966).

The problem of stimulus detection examined in the light of statistical decision theory is analogous to our examination of the problem of stimulus adaptation in the light of statistical estimation theory. Just as in the theory of ideal observers, it is not claimed here that the human observer is in any sense an optimal estimator of contrast; but rather that comparisons between the statistically optimum processor and human performance may, as in the detection/decision theory case, reveal facts about the structure of the signal processing mechanism in the visual system when it deviates from the optimum response. We present here the first few steps in applying estimation theory to the stimulus adaptation problem by assuming that the ideal and human observer are, in the crudest first approximation, identical. Many experiments exploring the precise relation between ideal and real observers are suggested by this approach but we shall leave these to remain in the future.

One may be interested to note that in practice, a simple numerical algorithm (steepest descent) for solution of Equations 7.3.4 exhibits a convergence behavior remarkably similar to the time course of induction of the adaptation effect known to exist in the visual system (Blakemore and Campbell 1969). It is tempting to speculate on the possible connections between two processes that produce such similar behavior. For the

actual solution of the ML equation we have used a modified Newton-like algorithm called directional discrimination (with Farris-Law Rule) to overcome the possible indefiniteness of the Hessian, cf. Bard (1974).

7.4 The Complete Model

In this section we shall present the complete proposed model of the brightness perception mechanism of the human visual system and discuss its performance with respect to the phenomena outlined in Chapter 6. Before we can present the complete model, however, we need to show how the ML contrast estimators are arrayed in the visual field. The previous sections have dealt with only a single idealized receptive field; we must first show how these receptive fields combine to perform the contrast estimation function necessary for the adaptation effect.

The reader may have noticed that Equations 7.3.4 contain an integral term with improper limits. Keeping the discussion of Section 5.5 in mind, we see that the proper way to obtain finite bounds for this integral involves windowing the integrand by a function dependent on . Thus the Equations 7.3.4 become:

$$(7.4.1a)\quad \int_{-\infty}^{\infty} h(\omega(x-\xi))\{r(\xi)-\log L_0(1+m\sin\omega\xi+\phi)\}\frac{\sin\omega\xi+\phi}{1+m\sin\omega\xi+\phi}\,d\xi=0$$

$$(7.4.1b)\quad \int_{-\infty}^{\infty} h(\omega(x-\xi))\{r(\xi)-\log L_0(1+m\sin\omega\xi+\phi)\}\frac{m\cos\omega\xi+\phi}{1+m\sin\omega\xi+\phi}\,d\xi=0$$

where h is some suitably chosen window function. The equation is solved for m and ϕ for fixed ω,x to obtain functions

$$\hat{m}=\hat{m}(\omega,x)$$
$$\hat{\phi}=\hat{\phi}(\omega,x)\,.$$

Thus for each point x in the visual field and each spatial frequency ω there is an estimate $\hat{m}(\omega,x)$, $\hat{\phi}(\omega,x)$ of the contrast and phase of the incoming signal. The links between this formulation and the receptive field/spatial filter model are strong in that the estimator reduces to a correlator in the limit for low contrasts. Note also that the $\hat{m},\hat{\phi}$ estimates are strong functions of x, thus indicating a model capable of localized response. This property plays a decisive role in the success of the full model in explaining the Tolhurst effect as well as the phenomena reported by Legge. This will be discussed more fully below.

The entire model is displayed in Figure 7.9. This model bears a resemblance to the receptive field/spatial filter model. Indeed, with a few exceptions, it is an extension of the latter to cover the adaptation response of the visual system. In Chapter 5, we showed that the Mandala transform could be considered essentially to be a bank of bandpass filters. However, the converse is not necessarily valid; not every bank of bandpass filters can be expressed as a Mandala transform. The symmetry group constrains not only the type of filter kernel in each channel but also the manner in which the kernel varies from channel to channel. Thus the Mandala transform is a

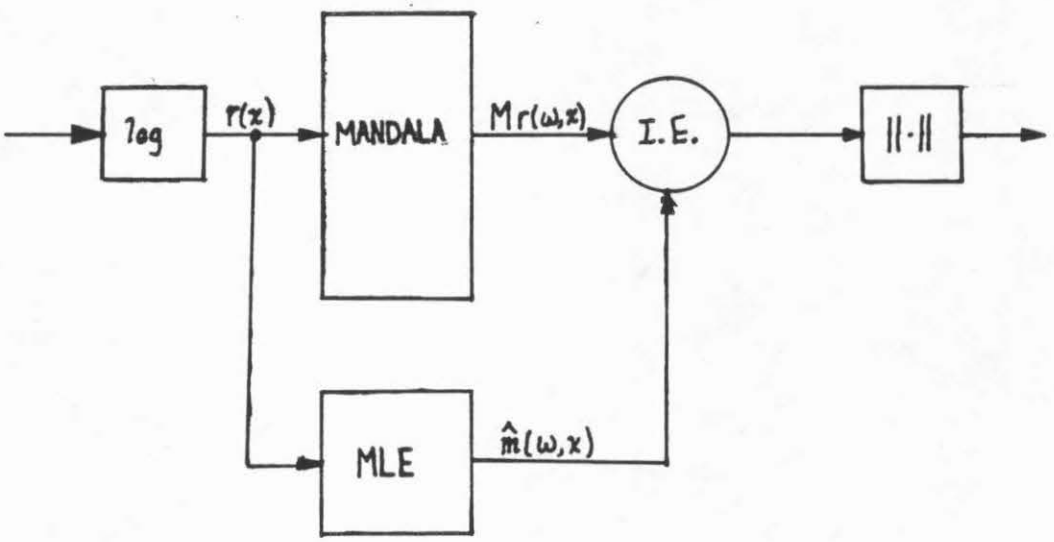


Figure 7.9 - Entire model.

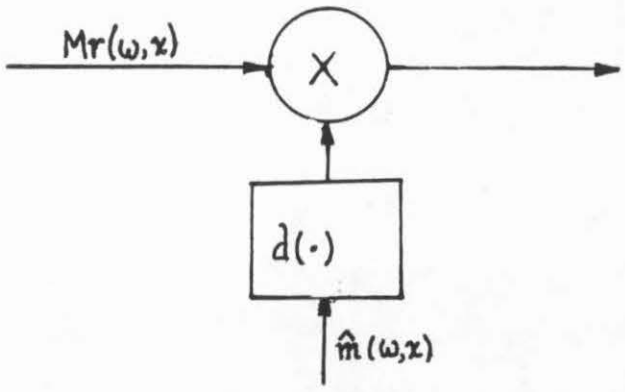


Figure 7.10 - Interaction element.

restricted receptive field/spatial filter model in which the frequency responses of each channel are identical except for scale and location of center frequency -- the former being proportional to the latter. Also note that in Chapter 5 we have derived that the bandpass characteristic should be symmetric about the center frequency (on a linear frequency axis). This condition is not met by most of the proposed receptive field/spatial filter models.

Recall that the Mandala transform required the choice of a window function. The shape of the window function is not too important in practice. Usually one chooses the function in keeping with desirable requirements such as maintenance of the frequency domain characteristics of the original waveform, cf. Blackman and Tukey (1958). We have chosen a Gaussian window.

Ignoring the interaction of the Mandala transform output with the adaptation stage for a moment we next treat the threshold stage. In this processing stage the signal is simply thresholded against some predetermined level common to all channels. The familiar variation in threshold among sinusoidal gratings of differing spatial frequencies is accomplished by the frequency response of the lateral inhibitory stage preceding the Mandala transform.

The adaptation mechanism is composed of a contrast

estimator followed by an interaction element. The contrast estimator has been developed in Section 5.3 and the initial part of this section. It is a processor that accepts an input signal r(x) and produces a signal $\hat{m}(\omega,X)$ which is the solution to Equations 7.4.1.

As far as the interaction element is concerned, matters are not nearly so clear. Again we are left without a rich enough group to allow use of the formal modelling approach. This is due to the paucity of available data on suprathreshold response characteristics of an adapted spatial frequency channel. Here again, unfortunately, we must proceed in an ad hoc manner. Figure 7.10 shows the details of the proposed interaction element. The contrast estimate $\hat{m}$ is passed through a memoryless nonlinear transformation $d(\cdot)$ to produce a signal $d(\hat{m})$ that multiplies the output of a given channel. The nonlinearity $d(\cdot)$ is necessary to produce adaptation behavior observed in various experiments. Adaptation raises the detection threshold by a ratio roughly proportional to the log of the adapting threshold for low adapting contrasts. At higher contrasts the adaptation effect saturates at a threshold elevation ratio of approximately 4 dB. See Tolhurst (1972) for more details. The memoryless nonlinear transformation $d(\cdot)$ is necessary to model this saturation behavior.

We will now discuss the performance of the complete

model with respect to the stimuli presented in Chapter 6.

First, we check that the estimator m(,x) is frequency specific. This is necessary because the estimator is connected to the input and not the output of the filter bank. Figure 7.11 displays the frequency specificity for the estimator at =1, x=0. Note that the frequency specificity of the estimator in this model is not tied to the bandwidth of the channels in the filter bank. Thus the narrowband/wideband channel controversy has a relatively simple solution in terms of this model -- simply let the estimator frequency specificity be much less tight than the bandwidth of the filter bank channels.

We now come to the first of the visual illusions, the Mach band illusion. Because the first stages of this model are very similar to the homomorphic lateral inhibition model, the mechanism generating the Mach illusion is the same in both models. The principal differences between these models occur when the contrast estimator produces an estimate that, through the interaction element, perturbs the output of a channel signal significantly. Roughly this occurs when a high amplitude frequency component of the stimulus is present. Figure 7.12 displays the homomorphic-lateral inhibition model response to a step-linear ramp stimulus, similar to Figure 6.2. Note that although the Mach bands occur for the ramp, by far the largest Mach response occurs at the

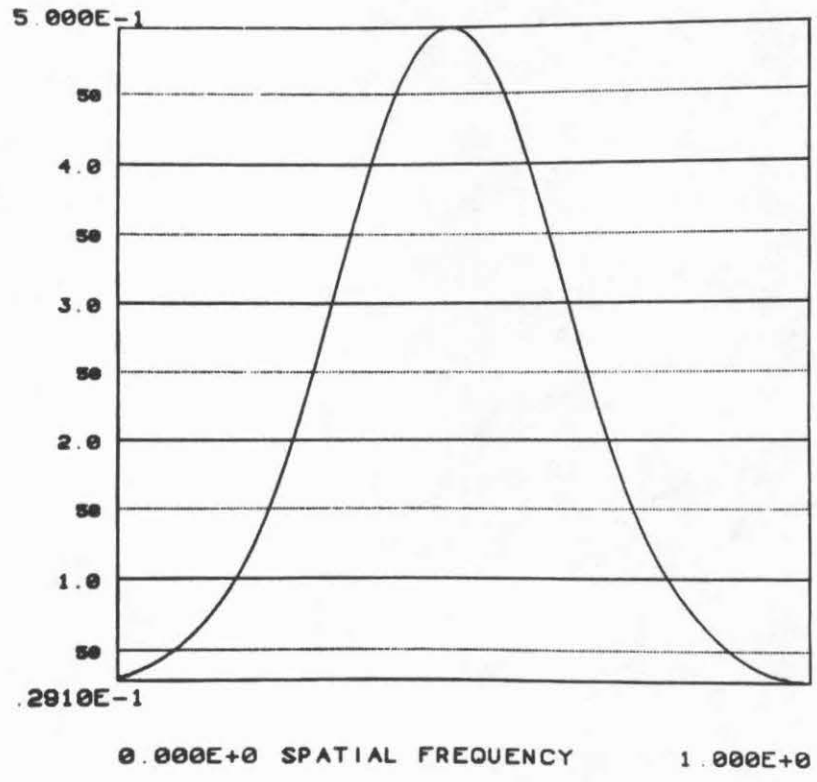


Figure 7.11 - Frequency specificity of ML estimator.

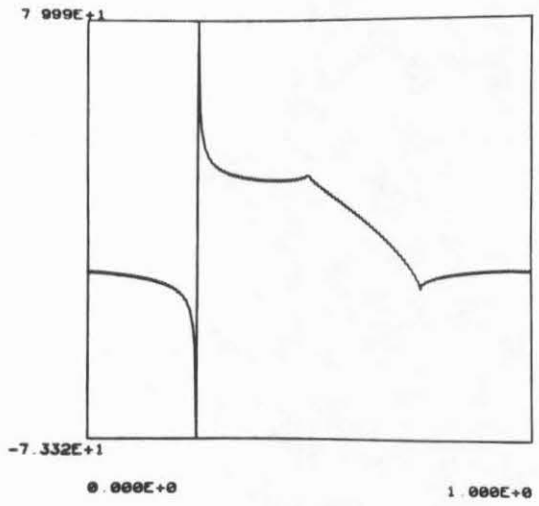


Figure 7.12 - Homomorphic step-ramp response.

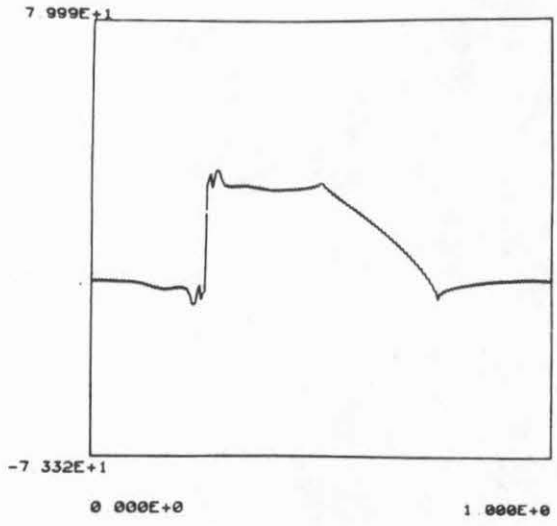


Figure 7.13 - New model step-ramp response.

step -- in disagreement with responses elicited from the visual system. Figure 7.13 shows the reponse of the new model to the same stimulus. Here, the Mach bands at the ramp remain relatively unchanged while the bands at the step are severely diminished. We have not, however, been able to cause the Mach bands to completely disappear at the edge. Thus we must conclude that either the Mach band is suppressed by some additional mechanism or that the chosen interaction element, in spite of its adequacy in predicting the threshold shift, does not quite model the actual suprathreshold interaction between contrast estimate and channel signal. (There is some experimental evidence for the latter possibility, see Blakemore, Muncey and Ridley (1973)). Answers to these questions will be left to some future second-generation model.

The new Mach softening phenomena are predicted (qualitatively) by this model. Because the contrast estimator derives its estimate from a portion of the signal, namely the portion under the window $h(\omega(x-\xi))$ (see Equation 7.4.1), the strip containing the luminance ramp is affected by the surrounding high contrast edges through the adaptation mechanism driven by the contrast estimator. Furthermore, low frequency estimates are gathered from larger areas so that the adaptation mechanism softens the low frequency components of the Mach band more effectively than the high frequency components. This is why softening

the Mach band makes it appear thinner. Varying the viewing distance of Figure 6.22 shows that this explanation also predicts the widening of the softened Mach band as the distance is lessened, i.e. increasing the angle subtended by the strip. This causes the principal adaptation effect to occur at a lowered spatial frequency thus allowing the original adapted spatial frequencies to participate in the Mach illusion mechanism, thus widening the band. Conversely, when the distance is increased, the softened bands should (and do) become much thinner. Figure 7.14 compares the responses of the homomorphic lateral inhibition model to the new model for a square wave train stimulus. On the left is shown the severe Mach banding that the homomorphic model predicts; the right shows how adaptation lessens this effect.

The second illusion we consider is the edge scalloping illusion. It is here we first use the thresholding properties of the model in conjunction with the unconscious inference idea. Recall that the explanation for the edge scalloping effect is quite similar to Campbell's explanation of the MF illusion. The lateral inhibition mechanism attenuates the low frequency portions of a stepwedge stimulus, producing a signal in which the relative magnitudes of the higher spatial frequencies cause the edges to "overshoot". When this signal is analyzed by a filter bank and each separate

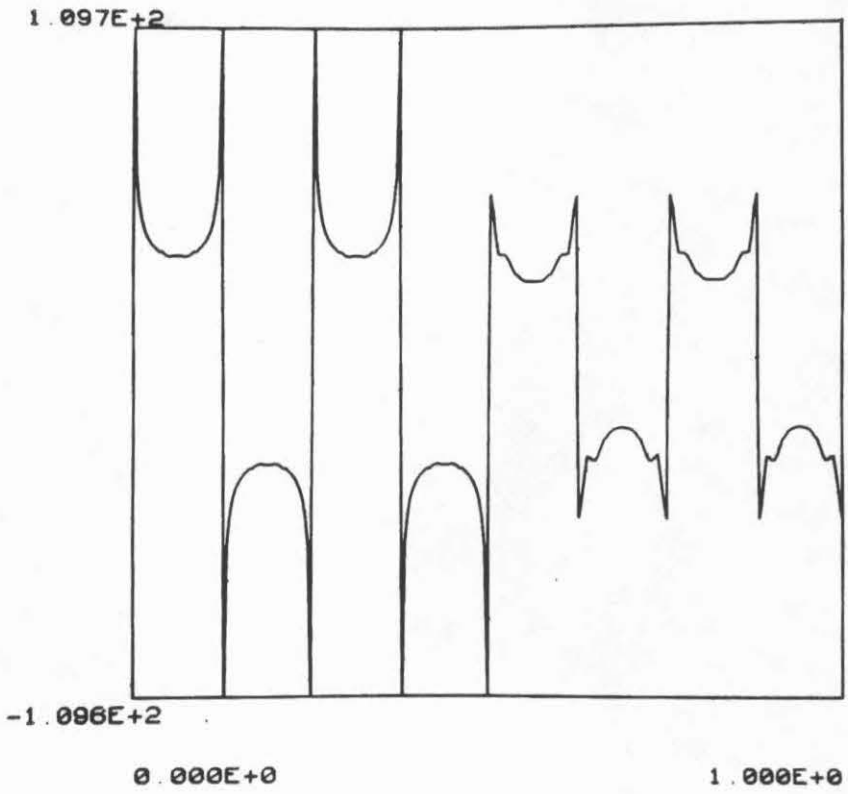


Figure 7.14 - Homomorphic square wave response vs. New model square wave response.

output channel thresholded, as in our model, it is possible the response produced will be similar to a signal that has not been passed through a lateral inhibition stage. Thus it would, under certain circumstances, be difficult for any mechanism in the visual system to decide whether or not the low frequencies in a signal have been attenuated. This is because the low frequency amplitudes of both attenuated and nonattenuated signals fall below detection threshold for the visual system. In our model an unconscious inference about the stimulus producing a given channel response shall be represented by the process shown in Figure 7.15, i.e. passing the channel response through an inverse Mandala transform followed by an inverse lateral inhibition filter and exponentiator.

Let us illustrate these processes by a few examples. Figure 7.16 shows a single bipartite edge, a single step, passed through the homomorphic lateral inhibition model. Figure 7.17 shows the same stimulus passed through the model and then through the unconscious inference stage. Not surprisingly, there is little scalloping here since the low frequency portion lies below threshold and is thus restored to its original amplitude by the unconscious inference step. Note that, in agreement with responses elicited from the visual system, scalloping occurs for a stepwedge, i.e. a series of steps. Figure 7.18 shows that when the low frequency portions of a stimulus exceed

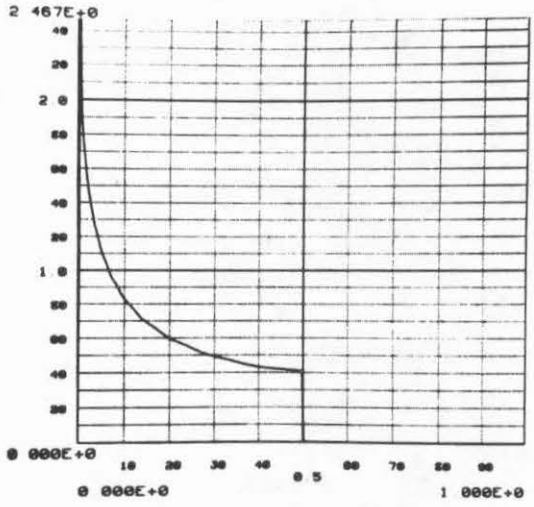


Figure 7.15 - Unconscious inference frequency response.

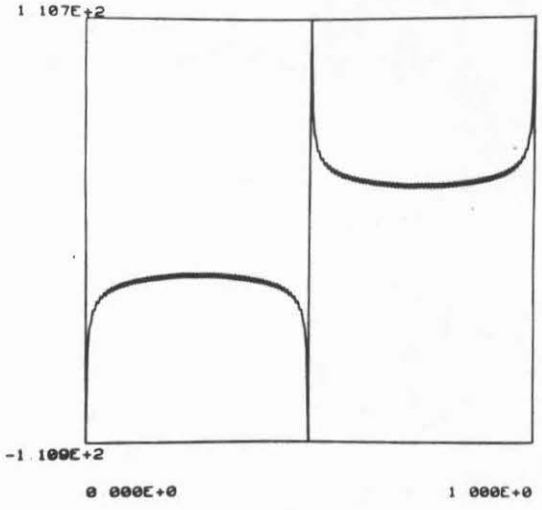


Figure 7.16 - Homomorphic edge response.

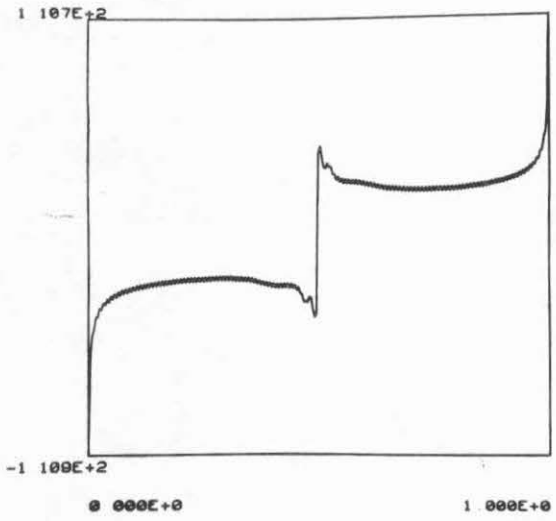


Figure 7.17 - New model edge response

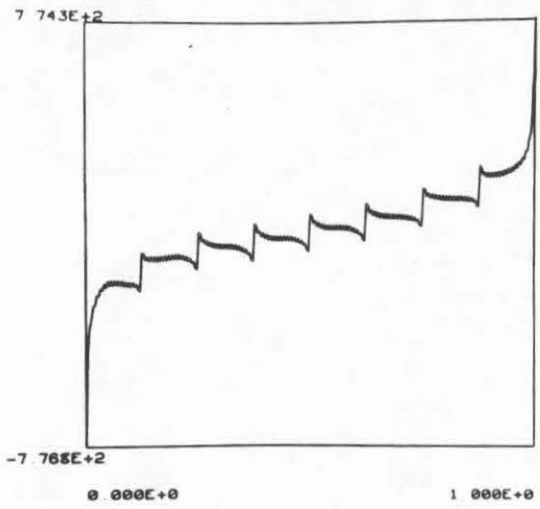


Figure 7.18 - New model stepwedge response.

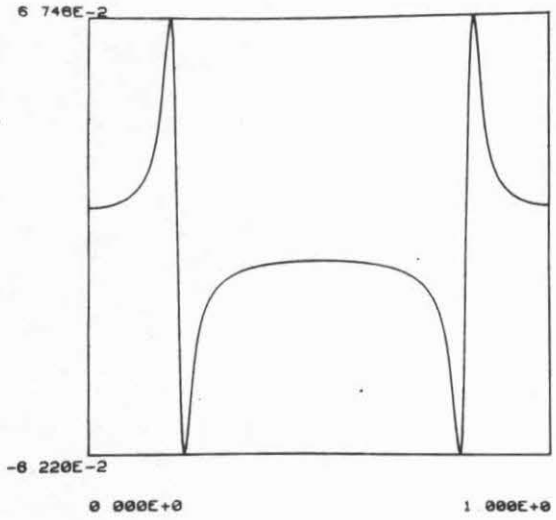


Figure 7.19 - New model Cornsweet response.

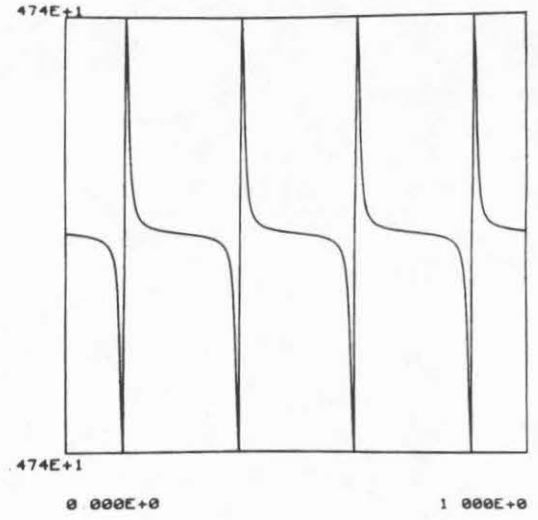


Figure 7.20 - New model multiple Cornsweet response.

the threshold of the system, it is then possible to infer that the low frequencies have been attenuated, thus scalloping occurs.

The third illusion we treat is the Cornsweet illusion. As we have discussed before, the multichannel detection model is able to explain the Cornsweet illusion through Campbell's unconscious inference explanation. We shall show here that with unconscious inference interpreted in the above manner, Campbell's argument not only is implementable on a computer but also produces the desired results. Figure 7.19 shows the unconscious inference of the model response to the Cornsweet illusion. The truly interesting behavior of this model compared with other models purporting to explain the Cornsweet response is revealed by presenting to the model a stimulus consisting of a series of Cornsweet edges. In contrast to the retinex based models (Horn's, Baxter's and Marr's) which predict that a stepwedge response should occur, our model correctly predicts that the Cornsweet response is softened by the addition of many other edges. This resonse is displayed in Figure 7.20. Note that this response is quite similar to the actual perceived brightness distribution. Finally, this model succeeds in correctly predicting that a Cornsweet response will occur when the stimulus is the new illusion of Chapter 6. This is shown in Figure 7.21. Again, in this case the retinex

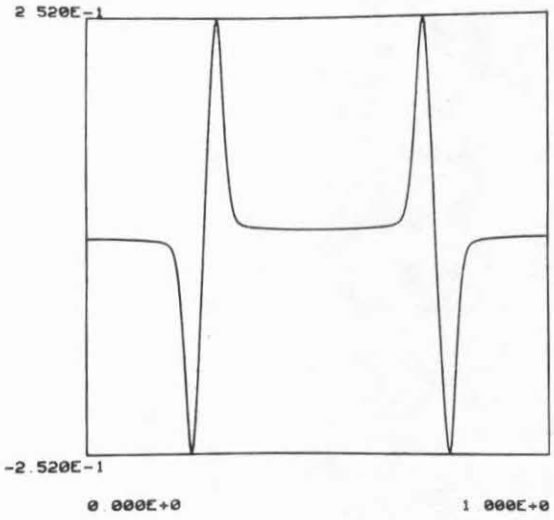


Figure 7.21 - New model New illusion response.

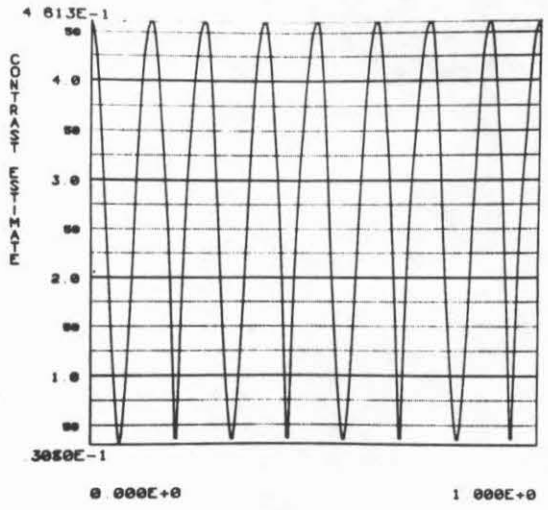


Figure 7.22 - Contrast estimate of third harmonic of square wave.

based models do not predict that the observed Cornsweet response occurs.

Finally, the model's performance with respect to the detection/adaptation data will be evaluated. Since this model encompasses the multichannel model which was designed to explain detection phenomena it is not surprising that the data in Sachs, Nachmias, and Robson (1971) are accounted for. Furthermore, because the model is specifically designed to include the adaptation response, it is apparent that this model accounts for the Blakemore and Campbell (1969) data, especially when one takes into consideration the interaction element we have chosen. The first real test for the model vis a vis detection/adaptation data occurs when we consider the Tolhurst (1972) data. Recall the data showed that a high contrast sinusoidal grating produces an adaptation effect much smaller when presented as a component of a square wave grating than when presented in isolation. Figure 7.22 displays the output of the contrast estimator $m(\omega_3, x)$ for ω_3, the third harmonic of the square wave input stimulus. The crucial point here is that the contrast estimate is <u>not</u> <u>uniform</u>: there exists a non-negligible spatial variation in m. This observation is the key to an explanation of the Tolhurst effect dispensing with the usual complicated scheme positing mutual inhibition of adaptation channels at high contrasts. Given a memoryless

interaction element -- its precise nature is not important -- we can see from Figure 7.22 that the spatial nonuniformity of the contrast estimate will adapt only certain portions of the third harmonic channel. Thus when the detection step of the adaptation experiment is performed portions of the channel will respond more vigorously than others. Depending upon the norming operation in the detection system, e.g. for the usual probability summation across space we have an L norm for p=4 to 6 (cf. Mostafavi and Sakrison (1976)), the nonresponding portions of the detection channel will not cause a rise in the detection threshold as much as would an attenuation of the entire channel which would occur in the case of the sinusoid presented in isolation as a stimulus. Thus the Tolhurst effect can be seen to be merely a phenomenon arising from non-global adaptation to non-global stimuli.

The data from Maudarbocus and Ruddock (1973) is consistent with our model since we were careful to design a nonlinear estimator that nullifies the effect of the receptor nonlinearity. Finally, the data from Legge is consistent with our model by an argument identical with the discussion of the Tolhurst effect. Indeed, in this case the phenomenon of non-global adaptation to local stimuli is quite transparent.

7.5 The Model as Image Processor

In this section we shall briefly explore the significance of our visual model to the image processing discipline. Second, we shall discuss the types of image metrics that the model implies. Third, we shall perform some elementary image processing algorithms based on the detection part of our model.

When one considers many algorithms for the coding, enhancement and restoration of images one finds that most of these algorithms optimize some global criterion. For example, in most coding schemes the mean square error is the criterion usually chosen to evaluate a given code in the class of codes with some prescribed structure constrained to fall below some transmission rate. Usually the ergodic assumption is invoked, thereby causing all statistical parameters such as sample means, sample autocovariances, etc. to be evaluated on a global basis. Again, much enhancement work is done either by collection of global statistics as in a histogram equalization approach or by filtering an image, which when considered in the frequency domain is a global concept. Finally, in restoration of images, the choice of metrics similar to those chosen by the coding discipline along with assumptions of spatial isotropy (symmetry under translations!) leads naturally to the power spectral density and through ergodicity to the frequency domain, a

global construct.

The primary objection to such global algorithms is that they are misguided. It is simply not the case that all regions of an image are equally important to the human visual system, e.g. certainly edges and small regions of high detail deserve more attention than either large, relatively flat areas or large areas containing much detail. Global algorithms expend much effort by allocating as much or more weight to the latter as to the former. For example, in coding, unimportant regions may be allocated precious codewords from a limited supply. Again, in enhancement, global algorithms permit unimportant regions to unduly influence the enhancement scheme through critical statistics. Finally, global algorithms sacrifice restoration accuracy in an image's important components for the sake of accuracy in relatively unimportant regions.

Nor are purely local algorithms, edge detection, curve detectors, etc. satisfactory. In fact, most researchers have found that the inclusion of texture into local schemes is an important addition. This recent flurry of activity in texture can be seen to be an attention to the (local) Fourier aspects of an image after long (over-)emphasis on the purely geometrical.

Our model and the Mandala transform provide a natural setting for such investigations. Additionally these have

a solid foundation in the symmetry theory of vision. There are global aspects in that each coefficient in the Mandala transform domain represents a Fourier-like frequency term. On the other hand there are local aspects in that a perturbation of the image in a region more than some number of periods away from the point at which the Mandala coefficient is being evaluated has no effect upon the value of that coefficient. Note that this is in contrast to almost all local schemes that are merely adaptations of some global process by segmenting the image into a series of procrustean "frames". We have already discussed how this windowing process destroys the symmetry group of vision replacing in its stead a most unnatural symmetry.

As for the model itself, there are indications that two algorithms fit into our scheme. The first is the homomorphic scheme, the second the maximum entropy restoration method. Both of these methods are closely related although their philosophies are quite different. The first proceeds from the physics of image formation, the second from the positivity of an image. Both methods exponentiate the input before performing either linear filtering or linear estimation on the resultant data. Thus, though these algorithms are primarily global, they share with the model the initial processing step, namely the receptor nonlinearity.

We now discuss how the model may be used to influence the choice of a metric for use in the processing of images (see Sakrison 1977). First of all, the receptor nonlinearity implies that any metric must first operate on the images with an exponential characteristic. This point is fully treated in Stockham (1972). Furthermore after a transformation by a receptor nonlinearity the model would have us take a Mandala transform of the resultant data. We are then left with an array of numbers which may be normed in some manner. The crucial property of this metric is the local/global characteristics inherited from the Mandala transform. Namely that 1) perturbations at moderately separated spatial regions (i.e. a number of periods) have independent effects upon the metric, 2) perturbations at moderately separated frequency bands (greater than one octave) also have independent effects upon the metric.

The qualitative effects of the threshold and adaptation mechanisms may be seen as primarily constraints upon the shape of the norming functional. The threshold element implies that small errors are just as good as no error at all. Thus it is useless to spend effort in reducing an error once it has reached a critical lower bound. The adaptation stage implies a similar behavior at the opposite end of the scale: because qualitatively the adaptation stage reduces the response to high contrast

stimuli that would have otherwise occured without adaptation, the effect of large errors is somewhat less than it would at first appear. Because of this we may say that very large errors are worse but not very much worse than large errors. Thus it is not effective to expend a great deal of effort to reduce very large errors if that effort does not significantly moderate such an error. To sum up: the picture of a norm that the model draws for us is as follows. At the low end of the contrast scale the norm must, below some critical value, fall to zero, i.e. the norm is very flat at the low end of the scale; correspondingly at the high end of the scale the norm must also flatten out. This picture is in marked contrast with the mean square norm used in practice. Although at the low end of the scale a mean square norm is not too different from our picture in that the square is relatively flat, at the opposite end of the scale, the type of qualitative behavior in the mean square norm is the opposite of what we desire. For larger and larger errors, a given increment of improvement carries more and more weight. This is in opposition to our picture which states that for large errors the increment of improvement shall diminish in importance. The effect that such a choice of norm has upon a processing scheme is important since much effort would be diverted from making improvements in moderate errors (which our model tells us

is the most important) toward making improvements in large errors (to which our model (and the visual system) is relatively insensitive).

We shall now present some crude examples of the use of our model in the processing of images. Unfortunately the computational requirements for the full two-dimensional algorithm lie approximately three orders of magnitude beyond capacity of the computer facilities available to the author (a DEC KA10) so that for the two-dimensional case only the detection portion of the model has been implemented.

The first image in Figure 7.23 shows the original image to be processed. First we pass the image through the receptor nonlinearity (a logarithm) and then perform the Mandala transform. The result is displayed in Figure 7.24. This picture is a representation of the Mandala transform output $|\hat{f}(\omega,x)|$. In the array of images each element is the portion of $|\hat{f}|$ corresponding to a fixed ω_o, i.e. a channel output. The array is arranged such that the first row corresponds to channels with center frequencies on a line through the $\omega_y=0$ axis. Each array element i,j (where i,j=-4,-3,...,0,1,...,4) represents a channel output $|f(\omega,x)|$ where

$$\|\omega\| = \omega_x^2+\omega_y^2 = \exp(i^2+j^2)$$

$$\angle\omega = \arctan(\omega_y/\omega_x) = \arctan(j/i).$$

A typical filter channel response is shown in Figure 7.25.

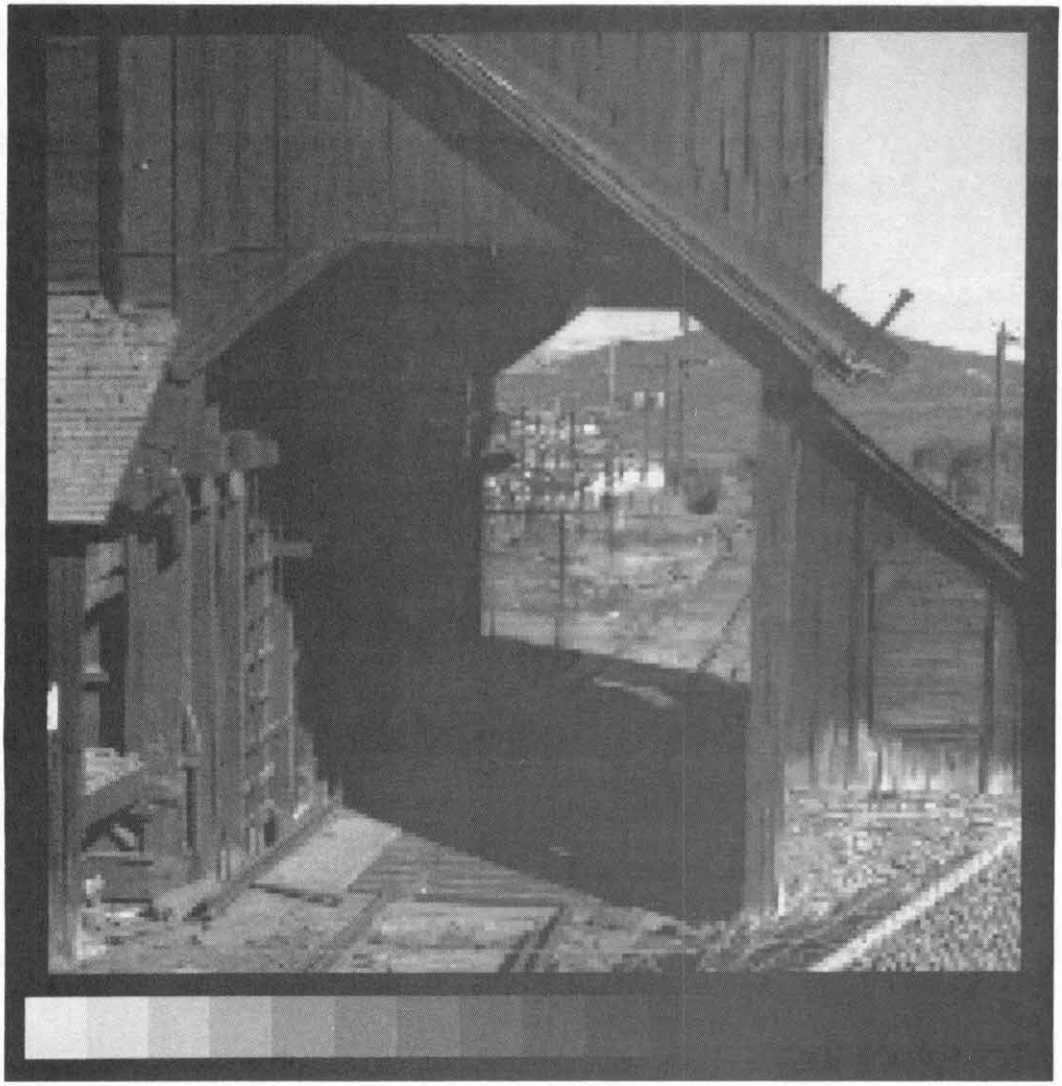


Figure 7.23 - Original "Mill".

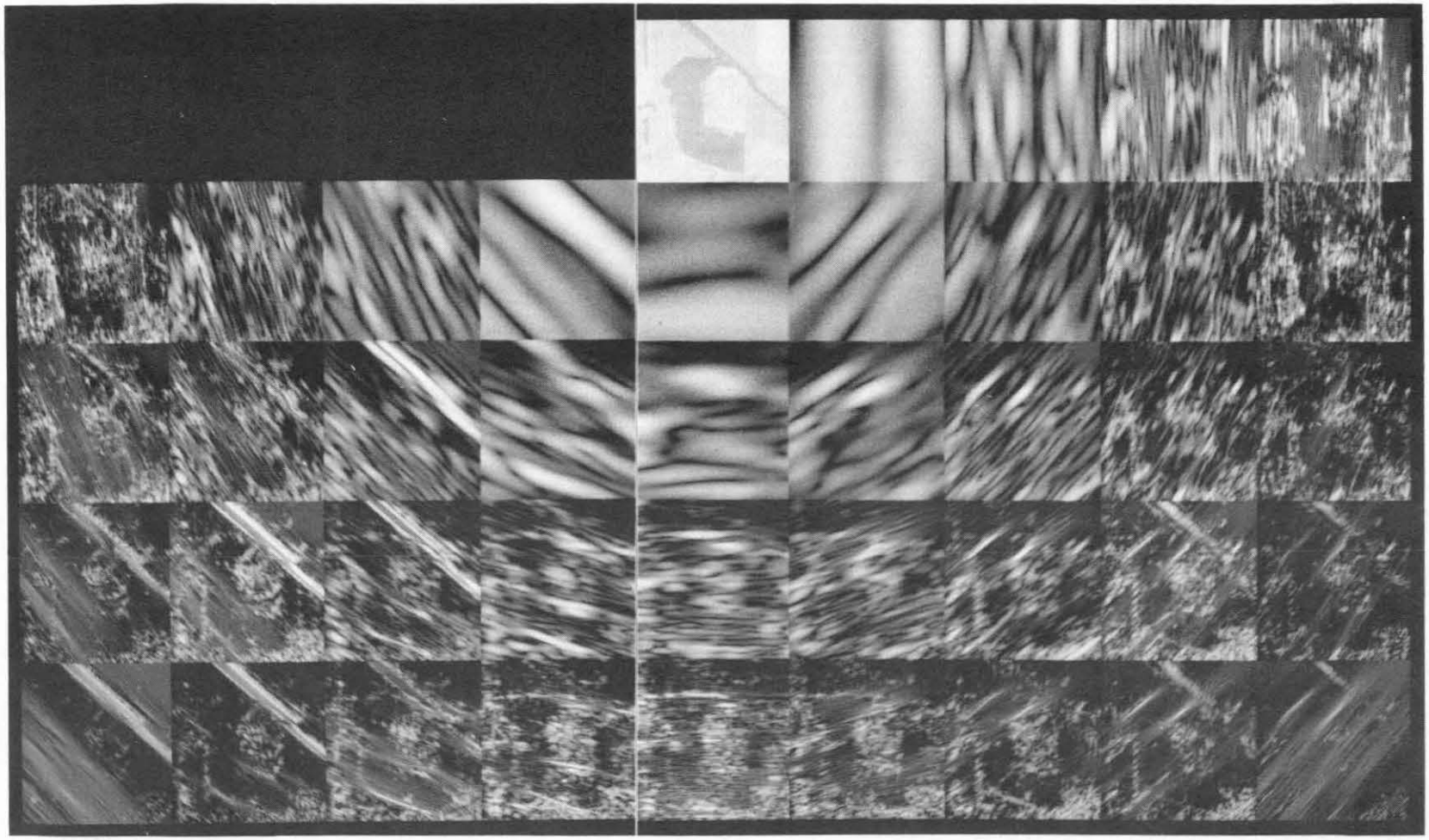


Figure 7.24 - Mandala transform of "Mill".

The frequency response is given by

$$k(\omega,\nu)=\exp(-Q_1\frac{\|\omega-\nu\|^2}{\|\omega\|^2})\exp(-Q_2\ A(\omega,\nu)^2\ )$$

where Q_1=3 and Q_2=33.2 and $A(\omega,\nu)$ is the angle between ω and ν given by

$$A(\omega,\nu)=\arccos(\frac{\langle\omega,\nu\rangle}{\|\omega\|\ \|\nu\|})$$

where the principal branch of arccos is taken. Since the channel outputs possess conjugate symmetry only half of the complete array is displayed.

To complete the detection portion of the model we scale each channel according to the modulation transfer function of the lateral inhibition stage. (Note that this is equivalent to preceding the Mandala transform by a lateral inhibition filter). Finally a threshold operation of each scaled channel is performed. In order to view the result we perform an inverse Mandala transform and exponentiation to obtain Figure 7.26. The obvious relationship of this procedure to the homomorphic enhancement procedure (Stockham 1972) behooves us to display the image processed homomorphically for comparison. This is done in Figure 7.27. Not surprisingly, the two enhanced images are quite similar in that they illustrate the well-known enhancement of detail in the low-illuminance portions of the image. Note however that the image produced by the new processing algorithm has an overall cleaner aspect. In fact, one is quite aware that there has been a removal of some image

```
MIN=   0000000      (   0000000      )
MAX=   9997470      (   9997470      )
AVGE=   5027561@-1
```

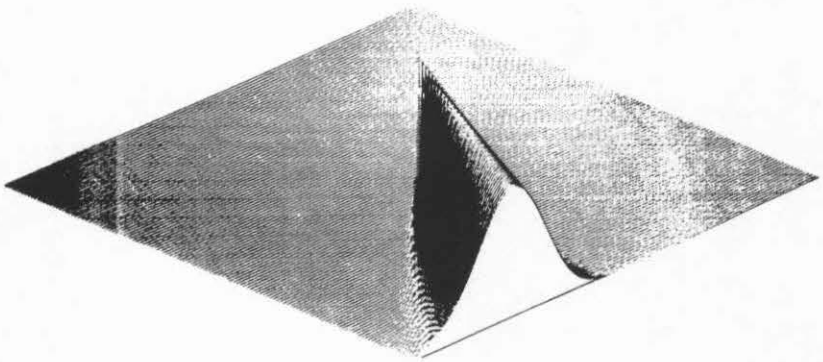


FIG7 25 SUBFILE A

Figure 7.25 - Filter channel frequency response.

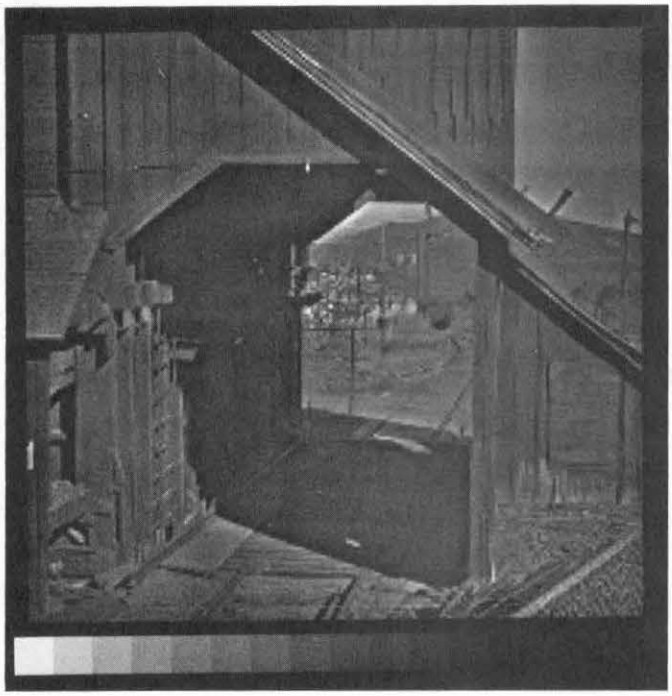


Figure 7.26 - Model response enhancement.

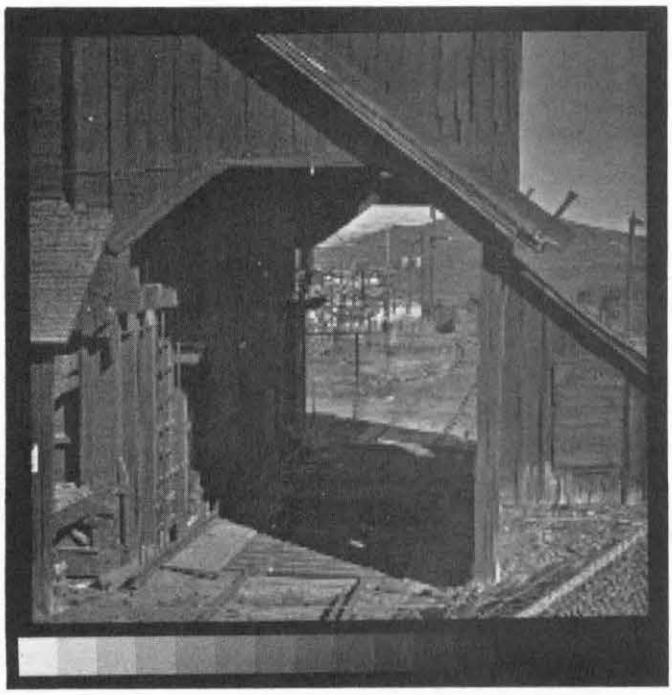


Figure 7.27 - Homomorphic enhancement.

features by the new processing algorithm but that the perceptually important aspects of that image have been retained. Note that thresholding the channel signals is a form of noise removal very similar to certain techniques in the speech processing community (Boll 1977).

Because this model transforms an image into a representation that is more "perceptual" (the Mandala transform domain), i.e. the components of this representation are decoupled, it should be possible to do kinds of nonlinear processing which mimic the operation of the visual system by performing operations similar to the visual system model. Recall that in the explanation for certain illusions the process of unconscious inference was crucial. We shall perform a similar operation for the two-dimensional case and show that it can be considered to be an operator that discards the texture and retains the form in an image. This is done by passing the image through the model and reconstructing by unconscious inference, i.e. an inverse Mandala transform followed by an inverse lateral inhibition filter followed by exponentiation. Note that this process is exactly analogous to the one-dimensional process. We simplify this process somewhat by scaling the individual channel outputs before performing the inverse Mandala transform. This results in a processing step of scaling, thresholding, and scaling by the reciprocal value, which

is the same as thresholding according to a threshold value set by the scale factor that varies from channel to channel. Figure 7.28 shows the image processed by such an operator whose threshold value is high for high spatial frequencies and low for lower frequencies. Clearly the form in this has been retained while the texture has been removed. The removed texture is displayed in Figure 7.29.

While these efforts represent relatively crude processing algorithms, it is hoped that the reader has been given some idea of the nature of future, more sophisticated algorithms based upon the psychophysical model presented here.

Figure 7.28 - Form extraction.

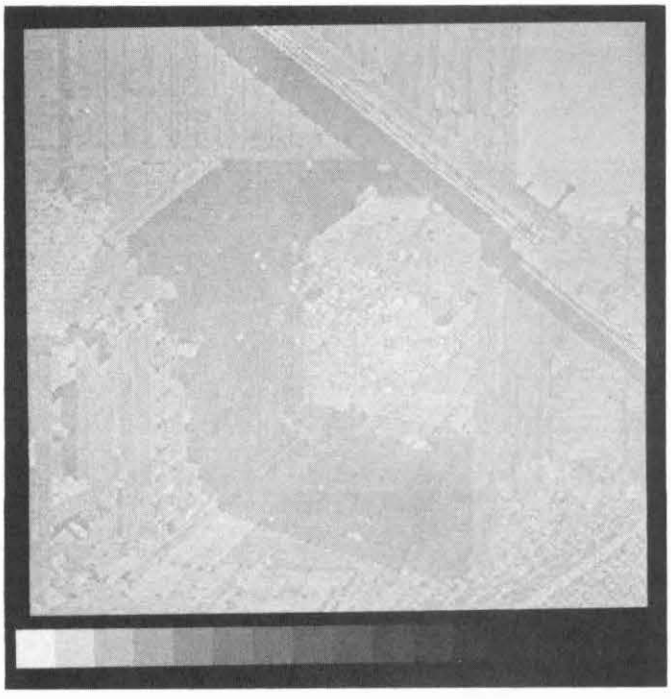


Figure 7.29 - Texture
Extraction.

CHAPTER 8

CONCLUSIONS

8.1 Summary

This thesis has sought to substantiate the claim that mathematical analysis properly has a much greater role in the study of perception than it has played in the past. We have taken the first steps toward what one may call a mathematical theory of perception. The key to open the application of mathematical analysis to the study of perception has been a repeated invocation of a single concept -- symmetry. The symmetry concept has allowed the introduction of the powerful methods of group representations.

We have developed a formalism that yields a structure formula for linear models of perceptual systems symmetric under an experimentally determined group of symmetry transformations. This structure formula is important since it avoids the practice of ad hoc modelling, viz. the positing of some arbitrary model structure along with successive modifications of the model to encompass the data of a sequence of experiments. In Chapter 6 we have seen that ad hoc modelling can lead to accretion of the

original model until it becomes so complex that the predictive utility of the final model is nil.

A symmetry theory for nonlinear models has also been developed. This theory assumes that the symmetry group actions on the input and output manifolds are known, along with a very general a priori form for the contemplated model. Here, ad hoc modelling is avoided because the a priori structure is of such a general structure, chosen once, and symmetrized to a special case. Essential use of the rudiments of infinite dimensional differential geometry is made in this theory.

The Mandala transform, a new transform developed in this thesis, is a result of the symmetry theories applied to the visual perception system. Even though this operator structure follows from only fundamental symmetry observations, the Mandala transform bears a striking resemblance to established psychophysical models and electrophysiological structures found in the visual cortex of various vertibrates. Furthermore, we have proposed that the Mandala transform is the proper effective algorithm approximating the Fourier transform and that it should replace the Short-time Spectrum in applications that process signals intended to be consumed by systems symmetric under shifts and dilatations.

Several new visual phenomena have been presented which shed new light on some old phenomena and which pose

challenges to certain established models. The principle of illusion softening has been applied to the classical Mach band phenomenon and new behavior of the visual system with respect to the Mach response has been found. Also, a new illusion eliciting the Cornsweet response without the use of sharp contours has been presented. Finally, adaptation of the visual system to a certain stimulus has been shown to be quite the opposite of what the most popular adaptation model would predict.

In the previous chapter, detailed calculations were presented that question the credibility of current extensions of the well-established multichannel detection model in that they largely ignore the effects of a receptor nonlinearity, an approximation valid for detection but not for adaptation studies. A new adaptation model using a nonlinear estimation technique is proposed. We have thus introduced the use of statistical estimation theory in adaptation studies. We have shown how this model agrees (sometimes only qualitatively) with known behavior of the visual system.

In developing this model we have given a new interpretation of unconscious inference (as an inverse filter-exponentiator) in terms of our model. Also, we have presented a simple explanation of the Tolhurst effect that requires no mutual inhibition between frequency channels. Finally, we have used the model as an image

processor and discussed, in general terms, the implications of our model for the image processing community at large.

8.2 Further Research

Possibilities for further research will no doubt suggest themselves to the reader when it is recalled that the symmetry theories do not in any way limit themselves to the situation studied here.

The application of symmetry groups to temporal models of human vision appear most attractive. In particular, for the study of adaptation phenomena to be brought completely under the method of formal modelling, we have pointed out the crucial nature of temporal considerations. Also, the study of space perception, that is, binocular vision, may be seen to be aided by diligent application of the symmetry theory using the more complicated groups of binocular space perception. This is particularly important since it seems that much of the system is devoted to the processing of these aspects of vision. Finally, the mechanism of form perception which remains one of the greatest scientific mysteries of today may be contemplated with these studies in mind.

As far as chromatic extensions of the proposed model are concerned, the work of Frei (1974) and Faugeras (1976) dovetail neatly with the present research. It would be

interesting to see a study combining the results of those researches with the present model.

The technique of psychophysical image coding as in Rom (1975) or Sakrison (1977) is based upon a given visual model. Implementations of these techniques have used the older visual models discussed above. Presumably, a more accurate visual model would allow these coding techniques greater freedom in choosing a code vector to be transmitted, thus allowing greater compression of the signal.

Another area of research which is most urgent with respect to some of the conclusions of this thesis is the proper psychophysical measurement of the various new phenomena presented in this thesis. For example, further study of the Mach softening response is necessary to completely accept or reject the hypothesis that the surrounding stimulus is masking or adapting those channels whose (large amplitude) response is necessary to the Mach illusion. Furthermore, it is most desirable to obtain precise measurements of the adaptation response to the near 100% duty cycle rectangular wave, because this response presents a significant challenge to the established Tolhurst model for adaptation.

Finally, applications of the symmetry theory to other perceptual systems, notably the auditory system, appear to have much promise. Currently at Utah there is activity

applying the Mandala transform to speech processing. Further work of modelling the auditory system itself has yet to be done and could lead to new and powerful techniques for the processing of sound.

REFERENCES

Alpern M., Rushton W.A. and Torii S. (1970) "The Size of Rod Signals" J. Physiol. Lond., v. 206, pp. 193-208.

Arveson W.B. (1976) Invitation to C^*-algebra, Springer-Verlag, New York.

Bard Y. (1974) Nonlinear Parameter Estimation, Academic Press, New York.

Baxter B. (1975) "Image Processing in the Human Visual System" Computer Science Tech. Rep. UTEC-CSc-75-168, Univ. of Utah.

v. Bekesy G. (1968) "Mach- and Hering-type lateral inhibition in vision" Vision Res., v. 8, pp. 1483-1499.

Blackman R.B. and Tukey J.W. (1958) The Measurement of Power Spectra, Dover, New York.

Blakemore C. and Campbell F.W. (1969) "On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images" J. Physiol. Lond., v. 203, pp. 237-60.

Blakemore C., Muncey J.P. and Ridley R.M. (1973) "Stimulus specificity in the human visual system" Vision Res., v. 13, pp. 1915-1931.

Boll S.F. (1977) "Application of the SABER method for improved spectral analysis of noisy speech" Computer Science Tech. Memo. UUCS-77-107, Univ. of Utah.

Burton G.S. (1973) "Evidence for non-linear response processes in the human visual system from measurements on the thresholds of spatial beat frequencies" Vision Res., v. 13, pp. 1211-1225.

Callahan M.W. (1976) "Acoustic signal processing based on the short-time spectrum" Computer Science Tech. Rep. UTEC-CSc-76-209, Univ. of Utah.

Campbell F.W., Cleland B.G., Cooper G.F. and

Enroth-Cugell C. (1968) "The angular selectivity of visual cortical cells to moving gratings" J. Physiol. Lond., v. 198, pp. 237-250.

Campbell F.W., Howell E.R. and Robson J.G. (1971) "The appearance of gratings with and without the fundamental Fourier component" J. Physiol. Lond., v. 217, pp. 17-18P.

Campbell F.W. and Robson J.G. (1968) "Application of Fourier analysis to the visibility of gratings" J. Physiol. Lond., v. 197, pp. 551-566.

Cassirer E. (1944) "The concept of group and the theory of perception" Phil. and Phenomonological Res., v. V, pp. 1-35. Also in Perception: Selected Readings in Science and Phenomenology, ed. Paul Tibbetts, Quadrangle/NY Times Book Co., New York, 1969.

Colas-Baudelaire P. (1973) "Digital picture processing and psychophysics: a study of brightness perception" Computer Science Tech. Rep. UTEC-CSc-74-025, Univ. of Utah.

Cornsweet T. (1970) Visual Perception, Academic Press, New York.

Dunn J.C. (1973) "Continuous group averaging and pattern classification problems" SIAM J. Comp., v. 2, pp. 253-72.

_____ (1975) "Group averaged linear transforms that detect corners and edges" IEEE Trans. Comp., v. C-24 no. 12, pp. 1191-1201.

Davidson M. (1968) "Perturbation approach to spatial brightness interaction in human vision" J. Optical Soc. Am., v. 58 no. 9, pp. 1300-1308.

Davidson M. and Whiteside J.A. "Human brightness perception near sharp contours" J. Optical Soc. Am, v. 61 no. 1, pp. 530-536.

Faugeras O.D. (1976) "Digital color image processing and psychophysics within the framework of a human vision model" Computer Science Tech. Rep. UTEC-CSc-77-029.

Frei W. (1974) "A new model of color vision and some practical implications" USCEE Rep. 430, pp. 128-143.

Fry G.A. (1948) "Mechanisms subserving simultaneous

brightness contrast" Amer. J. Optometry and Archives of the Amer. Acad. Optometry, v. 25, pp. 162-178.

Graham N. (1977) "Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels" Vision Res., v. 17, pp. 637-652.

Graham N. and Nachmias J. (1971) "Detection of grating patterns containing two spatial frequencies: a comparison of single channel and multiple channel models" Vision Res., v. 11, pp. 251-259.

Graupe D. (1972) Identification of Systems, D. Van Nostrand, New York.

Green D.M. and Swets J.A. (1966) Signal Detection Theory and Psychophysics, Wiley, New York.

Halmos P. (1950) Measure Theory D. Van Nostrand, Princeton.

Hille E. and Phillips R.S. (1957) Functional Analysis and Semigroups, Amer. Math. Soc. Colloquium Publ. 31, Providence, R.I.

Hoffmann W.C. (1966) "The Lie algebra of visual perception" J. Math. Psychol., v. 3, pp. 65-98.

_____ (1971) "Visual illusions of an angle as an application of Lie transformation groups" SIAM Review, V. 13 no. 2, pp. 169-183.

Horn B.K.P. (1973) "On lightness" MIT AI Tech. Memo. no. 295.

Hubel D.H. and Wiesel T.N. (1962) "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex" J. Physiol Lond., v. 160, pp. 106-154.

Julesz B. (1971) Foundations of Cyclopean Perception, Univ. of Chicago.

Land E.H. (1964) "The retinex" Amer. Scientist, v. 52, pp. 247-264.

Land E.H. and McCann J.J. (1971) "Lightness and the retinex theory" J. Optical Soc. Am., v. 61, pp. 1-11.

Landau H.J. and Pollack H.O. (1961) "Prolate spheroidal

wave functions, Fourier analysis and uncertainty -- II" Bell Sys. Tech. J., v. 40, pp. 65-85.

Lang S. (1962) Introduction to Differentiable Manifolds, Interscience, New York.

Legge G.E. (1976) "Adaptation to a spatial impulse: implications for Fourier transform models of visual processing" Vision Res., v. 16, pp. 1407-1418.

Loomis L.H. (1953) An Introduction to Abstract Harmonic Analysis, D. Van Nostrand, Princeton.

Mach E. (1865) "On the spatial distribution of the light stimulus on the retina", transl. F. Ratliff in Mach Bands: Qualitative Studies on Neural Networks in the Retina, Holden-Day, San Francisco, 1965.

_____ (1886) The Analysis of Sensations, transl. Dover, New York, 1959.

Mackey G.W. (1949) "Imprimitivity for representations of locally compact groups I" Proc. Nat. Acad. Sci. USA, v. 35, pp. 537-545.

_____ (1952) "Induced representations of locally compact groups I" Ann. of Math. (2), v. 55, pp. 101-139.

_____ (1953) "Induced representations of locally compact groups II" Ann. of Math. (2), v. 58, pp. 193-221.

_____ (1957) "Borel structures in groups and their duals" Trans. Amer. Math. Soc., v. 85, pp. 134-165.

_____ (1958) "Unitary representations of group extensions I" Acta Math., v. 99, pp. 265-311.

Marr D. (1974) "Computation of lightness in the primate retina" Vision Res., v. 14, pp. 1377-1388.

Maudarbocus A.Y. and Ruddock K.H. (1973) "Nonlinearity of visual signals in relation to shape sensitive adaptation responses" Vision Res., v. 13, pp. 1713-1737.

Maurin K. (1968) General Eigenfunction Expansions and Unitary Representations of Topological Groups, Panstwowe Wydawn. Naukowe, Warszawa.

Mautner F.I. (1950a) "Unitary representations of locally compact groups I" Ann. of Math. (2), v. 51, pp. 1-25.

_____ (1950b) "Unitary representations of locally compact groups II" Ann. of Math. (2), v. 52, pp. 528-556.

Mostafavi H. and Sakrison D.J. (1976) "Structure and properties of a single channel in the human visual system" Vision Res., v. 16, pp. 957-968.

Nachmias J., Sansbury R., Vassilev A. and Weber A. (1973) "Adaptation to square wave gratings: in search of the elusive third harmonic" Vision Res., v. 13, pp. 1335-1342.

Palais R. (1957) "A global formulation of the Lie theory of transformation groups" Memoirs of Amer. Math. Soc., no. 22.

Palm G. and Poggio T. (1977) "The Volterra representation and the Wiener expansion: validity and pitfalls" SIAM J. Appl. Math., v. 33 no. 2, pp. 195-216.

Pantle A. and Sekuler R. (1968) "Size-detecting mechanisms in human vision" Science, v. 162, pp. 1146-1148.

Pitts W. and McCulloch W.S. (1947) "How we perceive universals: the perception of auditory and visual forms" Bull. Math. Biophysics, v. 9, pp. 127-147.

Portnoff M.R. (1976) "Implementation of the digital phase vocoder using the fast Fourier transform" IEEE Trans. Acoust., Speech, and Sig. Proc., v. ASSP-24, pp. 243-246.

Quick R.G. (1974) "A vector magnitude model of contrast estimation" Kybernetik, v. 16, pp. 65-67.

Ratliff F. (1965) Mach Bands: Quantitative Studies on Neural Networks in the Retina, Holden-Day, San Francisco.

_____ (1971) "Contour and contrast" Proc. Amer. Phil. Soc., v. 115 no. 2, pp. 150-163.

Robson J.G. (1975) "Receptive fields: neural representation of the spatial and intensive attributes of the visual image" in Handbook of

Perception, v. 5, ed. E.C. Carterette and M.P. Friedman, Academic Press, New York.

Rock I. (1973) Orientation and Form, Academic Press, New York.

Rom R.J. (1975) "Image transmission and coding" Computer Science Tech Rep. UTEC-CSc-75-115, Univ. of Utah.

Sachs M.B., Nachmias J. and Robson J.G. (1971) "Spatial frequency channels in human vision" J. Optical Soc. Am., v. 61, pp. 1176-1186.

Sakrison D.J. (1977) "On the role of the observer and a distortion measure in image transmission" IEEE Trans. on Communications, v. COM-25, pp. 1251-1267.

Shannon C.E. (1949) "Communication in the presence of noise" Proc. IRE, v. 39, pp. 10-21.

Shepard R.N. and Metzler J. (1971) "Mental rotation of three-dimensional objects" Science, v. 171, pp. 701-703.

Stockham T.G. Jr. (1968) "Natural image compression with a quantitative error model" in Pertinent Concepts in Computer Graphics, ed. Fairman and Nievergelt, Univ. of Ill., Urbana; and "Intraframe encoding for monochrome images by means of a psychophysical model based on non-linear filtering of multiplied signals" Proc. of the 1969 Symp. on Picture Bandwidth Reduction, Gordon and Breach, New York, 1972.

_____ (1972) "Image processing in the context of a visual model" Proc. IEEE, v. 60, pp. 828-842.

Stockham T.G. Jr., Cannon T.M., and Ingebretsen R.B. (1975) "Blind deconvolution through digital signal processing" Proc. IEEE, v. 63, pp. 678-692.

Sullivan G.D., Georgeson M.A. and Oatley K. (1972) "Channels for spatial frequency selection and the detection of single bars by the human visual system" Vision Res., v. 12, pp. 383-394.

Sullivan G.D. and Georgeson M.A. (1977) "The missing fundamental illusion: variation of spatio-temporal characteristics with dark adaptation" Vision Res., v. 17, pp. 977-981.

Tolhurst D.J. (1972) "Adaptation to square wave gratings: inhibition between spatial frequency channels in the

human visual system" J. Physiol. Lond., v. 226, pp. 231-248.

van Nes F.L. and Bouman M.A. (1965) "The effects of wavelength and luminance on visual modulation transfer" Excerpta Medica, Int. Cong. Ser. no. 125, "Performance of the eye at low luminance" Proc. of the Symposium, Delft.

Weisstein N. and Bisaha J. (1972) "Gratings mask bars and bars mask gratings: visual frequency response to aperiodic stimuli" Science, v. 174, pp. 1047-1049.

VITA

Name	James Thomas Kajiya
Birthdate	July 13, 1951
Birthplace	Tokyo, Japan
High School	Magnolia High School Anaheim, California
College 1966-1968	Cypress College Cypress, California
University 1968-1972	University of Southern California Los Angeles, California
1975-1978	University of Utah Salt Lake City, Utah
Degree 1975	M.E., University of Utah Salt Lake City, Utah
Professional Positions	Project Engineer, Evans and Sutherland Computer Corporation, 1974-1975; Engineer, Quad-Eight Electronics, 1973-1974
Publications	with I.E. Sutherland and E.C. Cheadle (1975) "A random access video frame buffer" Proceedings of the Conference on Computer Graphics, Pattern Recognition and Data Structure, May 14-16, 1975, IEEE No. 75CH0981-1C
	(1976) "Group representations and the modelling of visual perception", in Image Science Mathematics, ed. C.O. Wilde and E. Barrett, North Hollywood

Western Periodicals Co.,
pp. 67-70

(1978) "The constant-Q
transform", in "Noise
suppression methods for
robust speech processing",
ed. S.F. Boll, Computer
Science Tech. Rep.
UTEC-CSc-78-073